DOCUMENT RESUME

ED 032 842 HE 001 132

The Development of a Computer Model for Projecting Statewide College Enrollments: A Preliminary Study

Rensselaer Research Corp., Troy, N.Y.
Pub Date Jan 68
Note-87p.; Report prepared for the Office of Planning in Higher Education, State Education Department, University of the State of New York, Albany, New York
EDRS Price MF-$0.50 HC-$4.45
Descriptors-*Computer Oriented Programs, *Cost Effectiveness, Educational Planning, *Enrollment Projections, Feasibility Studies, *Higher Education, Information Systems, Input Output Analysis, *Mathematical Models, Simulators
Identifiers-*New York

The purpose of this study was to develop the schema and methodology for the construction of a computerized mathematical model designed to project college and university enrollments in New York State and to meet the future increased demands of higher education planners. This preliminary report describes the main structure of the proposed computer simulation model as a non-stationary Markovian process with 3 basic segments: "initialization" (input of available data and setting the stage for system dynamics), operation (aging the system's student population and having it flow through the system), and output (describing the characteristics of students at different educational levels). This main structure is designed to act as a "calling" program for "contingency programs," which are separate computer models that provide data to educational planners and alter parameters in the main structure to show the effects of changes in the environment on enrollment projections. Features of the model would provide enough flexibility to meet growing informational needs and to use the capabilities of existing data systems in N.Y. State. Recommendations to the State Education Department of New York are (1) to construct a model that simulates student flows at 1 college as a prototype for a comprehensive state-wide model, (2) to develop a higher education information system for New York State, and (3) to conduct research on the development of desired functional relationships involved in contingency programs. (WM)

ED032842

THE DEVELOPMENT OF A COMPUTER MODEL FOR PROJECTING STATEWIDE COLLEGE ENROLLMENTS: A PRELIMINARY STUDY

Rensselaer Research Corporation
Troy, New York
January, 1968

This report was prepared for the Office of Planning in Higher Education, State Education Department, University of the State of New York.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HE 001 132

RRC

Rensselaer Research Corporation 1125 Peoples Avenue Troy, New York 12181 Telephone 518 / 270 - 6244

Dr. Ernest F. Nippes *President and Director*

March 12, 1968

Dr. Robert McCambridge
Assistant Commissioner in Higher
Education Planning
The State Education Department
Albany 1, New York

Dear Dr. McCambridge:

We are pleased to submit herewith our report on the development of a computer model for projecting statewide college enrollments.

It would be difficult to express appreciation to the many people who cooperated in this study. We are grateful to the Computer Model Advisory Committee, and the registrars and the planning officials at the institutions visited for their assistance in defining the higher educational environment in New York State. Members of both the planning and statistical services staff of the State Education Department have been very helpful in the formulation of the proposed computer model.

We shall look forward to the opportunity of discussing with you our findings and recommendations. We would be pleased to assist in any way possible during future discussions of the development of a computer model for projecting college enrollments in New York State.

Sincerely,

Ernest F. Nippes

Ernest F. Nippes
President

EFN:sg

TABLE OF CONTENTS

	Page
SUMMARY.	1
RECOMMENDATIONS.	4
SCOPE OF THE PRELIMINARY STUDY	5
1.1 Acknowledgements	7
HIGHER EDUCATION IN NEW YORK STATE: CHARACTERISTICS OF THE ENVIRONMENT	9
2.1 The Systems Approach to Higher Education . . .	9
2.2 Inputs to the Educational Process.	13
2.3 The Educational Activity	15
2.4 Outputs of the Educational Process	15
2.5 A Morphological Analysis	15
2.6 The Conceptual Model	21
PROPOSED METHODOLOGY FOR PROJECTING STATEWIDE COLLEGE ENROLLMENTS.	26
3.1 Introduction	26
3.2 An Overview of the Model	27
3.3 The Structure of the Model	31
3.4 The Projection Methodology	44
3.5 The Concept of Contingency Programs.	52
3.6 Summary.	55
FEASIBILITY OF PROPOSED METHODOLOGY.	56
4.1 Technical Feasibility.	56
4.2 The Information System: A Critical Factor	58
4.3 A Suggested Cost/Benefit Relationship.	62
APPENDIX A: INSTITUTIONS AND ORGANIZATIONS VISITED DURING THE COURSE OF THIS STUDY.	69
APPENDIX B: A SIMPLIFIED NUMERICAL EXAMPLE.	70
APPENDIX C: THE DESIGN OF A SAMPLING PLAN FOR ESTIMATING STUDENT FLOWS	80
APPENDIX D: COMPUTER MODEL ADVISORY COMMITTEE	83

LIST OF FIGURES

Page

Figure 1: SIMPLIFIED EDUCATIONAL SYSTEM 11

Figure 2: INPUTS TO HIGHER EDUCATIONAL PROCESS. 14

Figure 3: THE EDUCATIONAL ACTIVITY. 16

Figure 4: OUTPUTS OF THE EDUCATIONAL PROCESS. 17

Figure 5: ILLUSTRATION OF MORPHOLOGICAL ANALYSIS. . . . 19

Figure 6: MICRO STRUCTURE OF CONCEPUTAL MODEL 22

Figure 7: THE CONCEPTUAL MODEL. 24

Figure 8: FLOW CHART OF PROPOSED COMPUTER MODEL 45

Figure 9: FROSH-SOPH TRANSITION MATRIX, 1969. 76

Figure 10: SOPH-JUNIOR TRANSITION MATRIX, 1970 77

Figure 11: JUNIOR-SENIOR TRANSITION MATRIX, 1971 78

Figure 12: SENIOR-FIFTH-YEAR TRANSITION MATRIX, 1972. 79

Figure 13: SAMPLES SIZES TO ESTIMATE p WITH INDICATED PRECISION AND CONFIDENCE LEVELS 82

SUMMARY

1. Higher education in New York State can be visualized as a system consisting of an aggregate of micro-structural models. These models represent student flows either by curriculum, college (or type of college) or both. Associated with each micro-structural component are specified student inputs, and concomitant to these are certain student outputs — both characterizable in a number of ways. Upon identifying and delineating the variables relevant to a description of the system's input and output, it becomes apparent that the required information far exceeds its availability at the present time. Based on (1) interviews with registrars and planning officials at representative colleges and universities in the State; and (2) discussions with individuals actively engaged in educational planning for New York State, a feasible (rather than the complete) set of relevant variables were outlined and incorporated into a proposed model.

2. The main structure of the proposed computer simulation model is a non-stationary Markovian process; that is, simulated students are aged and flow through the modeled system in patterns governed by transition matrices whose elemental values are allowed to change over simulated time. The three basic segments of this process are initialization (input of available data and setting the stage for the system dynamics);

operation (aging the student population of the system and having it flow through the latter); and output (describing in detail the characteristics of students at different educational levels).

3. Using the building-block approach, the main structure of the proposed model is designed to act as a "calling" program for other sub-programs. The purpose of these sub-programs, called Contingency Programs, is providing information to educational planners to assist them in answering questions of the "what if" variety. Contingency programs are separate computer models which will systematically alter parameters in the main structure to show the effects of various changes in the environment on enrollment projections.

4. The reliability of the enrollment projections derived from the proposed computer simulation model are a direct function of the reliability, accuracy, and desired level of disaggregation of the input information. While the nature of the necessary data collection system is beyond the scope of this study, the high levels of disaggregation and sophistication of the proposed model allow for both flexibility and adaptivity to: (1) meet the increased informational needs of educational planners; and (2) utilize the growing capacity and capabilities of any data systems which might exist in New York State.

5. Detailed development of a Cost/Benefit relationship in the final section of the report aids in the delineation of the nature of the costs and benefits associated with the

construction and use of a computer simulation model for projecting statewide college enrollments. While the analysis may be used to evaluate successive stages of development of the statewide model, the derived relationship cannot be utilized until insights into the cost/benefit structure are gained through the development and testing of a prototype model.

RECOMMENDATIONS

1. Construction of a model simulating student flows at a single college — as a prototype to a more encompassing statewide enrollment projection model — should be begun. Insights will thus be gained into: (1) the costs and benefits of the larger model, and (2) the operating characteristics of a statewide data collection system amenable to use with a sophisticated computer projection model.

2. The development of a higher educational information system for New York State should begin as soon as possible. The full potential of the proposed projection model will not be realized until the input information is at the same level of sophistication as the model. Much of the form and substance of the data collected will be a function of the techniques used for projection of enrollments. Therefore, the data system should be developed in view of the requirements of the computer simulation.

3. Research must be started on the development of the desired functional relationships involved in the construction of contingency programs. This research must be carried on simultaneously with the construction of the enrollment projection model and the development of a higher education information system — otherwise, one or both of the latter may stand idle instead of making the important (and necessary) contribution of which they are capable.

SCOPE OF THE PRELIMINARY STUDY

The purpose of this study was to develop the schema and methodology necessary for the construction of a computerized mathematical model for projecting college and university enrollment in New York State. In addition, it was felt that the methodology should be designed in such a way that it will grow to meet increased demands of higher educational planners in the future.

The first stage of this study was the identification and delineation of the relevant variables that affect present college enrollment and those that may be assumed to affect future college enrollment in New York State. This phase was oriented toward developing a greater understanding of the environment in which any enrollment projection model must function. The activities included: (1) interviews with registrars and planning officials at representative colleges and universities in New York State; (2) discussions with individuals actively engaged in developing projections for educational planning within the New York State Education Department and the U.S. Office of Education; and (3) interviews with other knowledgeable people from organizations such as the American Association of Collegiate Registrars and Admissions Officers.[1] In addition, an extensive search and

---

[1] A listing of the organizations visited may be found in Appendix A.

review was made of the theoretical and applied literature on the subject of educational model building and enrollment projection. The results of these activities are presented in Section 2 of this report, HIGHER EDUCATION: CHARACTERISTICS OF THE ENVIRONMENT.

The final stage consisted of the development of the necessary methodology to construct a computerized model to project enrollment in the total higher educational system of the State. To accomplish this, the educational system was viewed as a sequential transformation process which accepts inputs from the college going population and provides desired outputs such as holders of specified degrees. A model of this transformation process was proposed and is presented in Section 3, PROPOSED METHODOLOGY FOR PROJECTING STATEWIDE COLLEGE ENROLLMENTS, with a simplified example described in Appendix B.

In addition, an evaluation was made of the feasibility of constructing a detailed predictive model for use in simulation procedures. Based on the knowledge that a continual trade-off must be made between model fidelity — i.e., how representative the model is of the real world — and costs of constructing and operating the model, this evaluation took into consideration such factors as quality and quantity of input data, methods of parameter estimation, and desired level of disaggregation of output information. This material, including a proposed cost/benefit relationship,

is presented in Section 4, FEASIBILITY OF PROPOSED METHODOLOGY; a discussion of possible sampling procedures is also given in Appendix C.

This study was conducted by an interdisciplinary team designed to take full advantage of the technical resources of Rensselaer Polytechnic Institute. The personnel who participated in this study with their area of specialization were:

A. Consultants:

Dean Bouton, Computer Science.

Roland T. Eustace, Ph.D., Educational Research and Public Finance.

Roman V. Tuason, Jr., Ph.D., Planning/Information Systems.

John W. Wilkinson, Ph.D., Statistics.

B. Research Associates:

Allen Baisuck, Ph.D. candidate, Operations Research and Statistics.

Russell C. Koza, Ph.D. candidate, Managerial Economics.

John J. Neuhauser, Ph.D. candidate, Operations Research & Statistics.

William A. Wallace, Ph.D., served as project director.

1.1 Acknowledgements

It would be difficult to express appreciation to the many people who cooperated in this study. Throughout the study activities, we found the tenor of the environment very conducive to realization of the goal of a computerized model

for projecting statewide college enrollments. All the individuals interviewed expressed a sincere desire to assist in any way possible in achieving this goal.

It must be noted, however, that without the close cooperation and guidance received from Dr. Robert McCambridge, Assistant Commissioner in Higher Education Planning and Mr. Thomas Shea, Associate Coordinator in Higher Education Planning, an applied research project of this nature could not have been accomplished; we express our sincere appreciation to them. In addition, the comments and suggestions of the Computer Model Advisory Committee (listed in Appendix D) proved invaluable in the preparation of this report.

HIGHER EDUCATION IN NEW YORK STATE: THE CHARACTERISTICS OF THE ENVIRONMENT

This section presents a description of the higher educational environment in New York State. Following a brief discussion of the systems approach, the characteristics of this environment are presented with a delineation of those characteristics which are amenable to analysis using the proposed methodology. In addition, a conceptual model of the educational system is developed which employs a modular structure permitting the building of basic models and combinations thereof to form an aggregate model describing the total educational system.

2.1 The Systems Approach to Higher Education

In order to adequately describe the complex nature of the educational process in New York State, it was felt desirable to employ a technique which recognizes the complexity and reduces the overall problem into manageable sub-sets which can be analyzed separately. Systems Analysis is a technique which provides a framework for visualizing these sub-sets with relevant internal and external environmental factors to form an integrated whole. Since most systems are dynamic rather than static, a method of reflecting change in the total environment and the associated effect on input/output relationships is necessary, and is provided in models utilizing the systems concept. Thus, this concept

is particularly applicable to problems involving the large complex systems typical of the educational environment.

Using a systems approach, an educational system may be viewed as a sequential transformation process which draws upon certain inputs to provide a predetermined output. Figure 1 shows this simple scheme. From the general population of the total outside environment, a certain segment can be delineated and designated as a college student population or college population pool. Contained within this pool are certain classes of individuals who may be admitted to a particular educational process for the express purpose of undergoing an educational transformation provided certain requirements are met. Within the educational activity itself, members undergoing this educational process are able to leave the system voluntarily or involuntarily if certain demands cannot be met. The end result of the educational transformation process is an output consisting of members of the total population who are now differentiated from the remainder of the population by virtue of certain characteristics resulting from exposure to this process.

Simply stated, a certain group of potential students may be selected from the general population for college entrance (or other forms of higher education). These same students have differing characteristics, such as origin, sex, or marital status. After entering the higher education process, these students may: (1) remain in school (provided they meet certain academic and administrative

SIMPLIFIED EDUCATIONAL SYSTEM

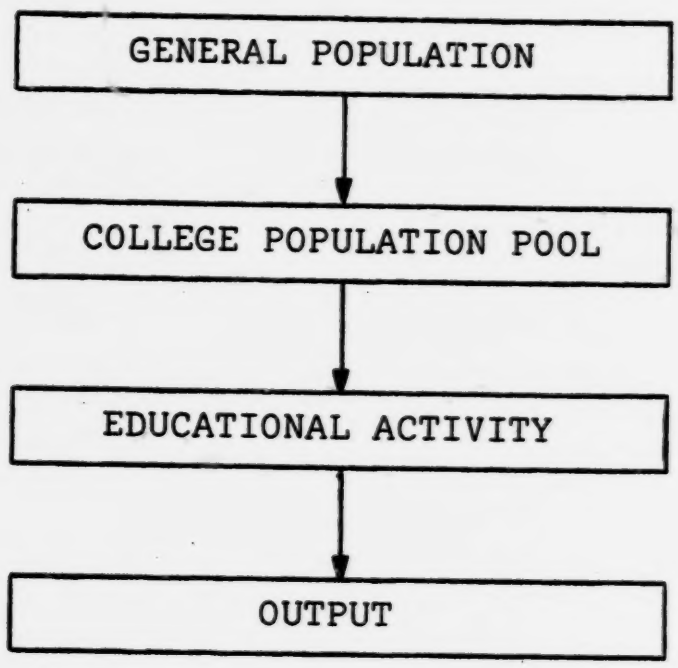


Figure 1

standards); (2) leave to enter the outside world, i.e., the work force, military service, etc.; (3) continue to the next level within the educational process. Moreover, they may elect to transfer to another educational curriculum within the same institution or to transfer to another educational environment, i.e., college, curriculum or both. Having met the system requirements and progressed through a prescribed curriculum, the student population is then able to meet the requirements for a specified output. As a result, the final output of the educational system is a sub-set of the general population possessing certain degrees (or certificates) conferred on the basis of satisfactory completion of requirements of the educational process. This output must also encompass those individuals whose educational objectives did not include the attainment of a degree or certificate but only the additional knowledge or skills which could be gleaned from one or more courses.

This description of an educational system is highly simplistic, but serves as a frame of reference for the development of a conceptual model. In order to proceed with the development of this model, it is next necessary to examine the relevant segments and classifications which exist in the input, educational process, and output of the educational system. A conceptual model can then be constructed which will have the capability of expressing system relationships in mathematical form (as discussed in

Section 3.3). This formulation permits manipulation for purposes of gaining insights into the dynamic process of higher education.

2.2 Inputs to the Educational Process

A possible delineation of the inputs to the higher educational process is shown in Figure 2. The college population pool is divided into segments based upon occupational status. Members of these segments can then be further classified according to a set of specified characteristics. It is expected that the college population pool will consist of:

(1) high school graduates;

(2) members of the work force who desire to return to school to improve their skills;

(3) members of the non-work force, e.g., homemakers and school dropouts not gainfully employed;

(4) people in the military service who may initiate or resume their educational activities.

These various segments of the college population pool may be classified with regard to age, sex, geographical origin, marital status, race, religion, etc., to further delineate the college population pool.

INPUTS TO HIGHER EDUCATIONAL PROCESS

SEGMENTS	CLASSIFICATIONS
HIGH SCHOOL GRADUATES	AGE
WORK FORCE	SEX
NON-WORK FORCE	GEOGRAPHICAL ORIGIN
MILITARY SERVICE	MARITAL STATUS
	RACE
	RELIGION

Figure 2

2.3 The Educational Activity

Certain inputs are drawn from the population according to standards imposed by the educational activity. Segments within the educational activity include undergraduate and graduate students and these groups may be further delineated as shown in Figure 3.

2.4 Outputs of the Educational Process

The final result of the educational process is the output of the system as previously described. The segments of this output are illustrated in Figure 4. Once again these segments can be classified further by a curriculum or college and also by the characteristics of the degree holder, i.e., age, sex and so forth.

2.5 A Morphological Analysis

As an aid to further analysis, the morphology or structure of the higher educational system can be shown schematically. This type of presentation depicts the magnitude of the dimensionality of the system. In addition, it shows how the dimensionality is related to the informational requirements of higher educational planners. As an example, consider the possibility that educational planners in New York State want to identify the geographical origin of individuals completing the educational process by receiving a degree in a specified curriculum, e.g.,

THE EDUCATIONAL ACTIVITY

SEGMENTS	CLASSIFICATIONS
UNDERGRADUATE	AGE
FRESHMAN	SEX
SOPHOMORE	ORIGIN
JUNIOR	MARITAL STATUS
SENIOR	COURSE LOAD
5th-YEAR	INSTITUTIONAL CONTROL
PROFESSIONAL	CURRICULUM
GRADUATE	COLLEGE/CAMPUS
MASTER'S	RACE
DOCTOR'S	RELIGION
POST DOCTORAL	

Figure 3

OUTPUTS OF THE EDUCATIONAL PROCESS

SEGMENTS	CLASSIFICATIONS
NON-DEGREE	AGE
CERTIFICATE	SEX
ASSOCIATE'S	ORIGIN
BACHELOR'S	MARITAL STATUS
PROFESSIONAL	CURRICULUM
MASTER'S	RACE
DOCTOR'S	RELIGION
POST-DOCTOR'S	

Figure 4

"number of biology majors from Nassau County who receive a Master's degree". Figure 5 shows the structure of the informational requirements, assuming only ten geographical origins (an acknowledged over-simplification), twenty-six curricula as specified by the Department of Health Education and Welfare, and seven degree designations.

Note that there are 10 X 26 X 7 = 1820 different cells. Some of these cells may be empty due to: (1) lack of current occupancy, e.g., no law majors receiving doctorates from origin #2, (say) Southern Tier; (2) non-existence of the category, e.g., no certificate programs in religion. This example is only one of many possible informational requirements. Using this type of analysis and the segments and classifications previously discussed, it is estimated that 4500 possible cells would be needed to describe the inputs and 60,000 data cells to describe the possible outputs of the educational system.

The size and complexity of any proposed methodology may now be seen to be a function of the number of categories (level of disaggregation) desired or specified. An analysis of the information and data collection requirements necessary to obtain each such level must then be matched against the importance attached to the attainment of such a level. This comparison will provide an answer in the form of feasibility or non-feasibility of such disaggregation. This subject is explored further in Section 4.

ILLUSTRATION OF MORPHOLOGICAL ANALYSIS

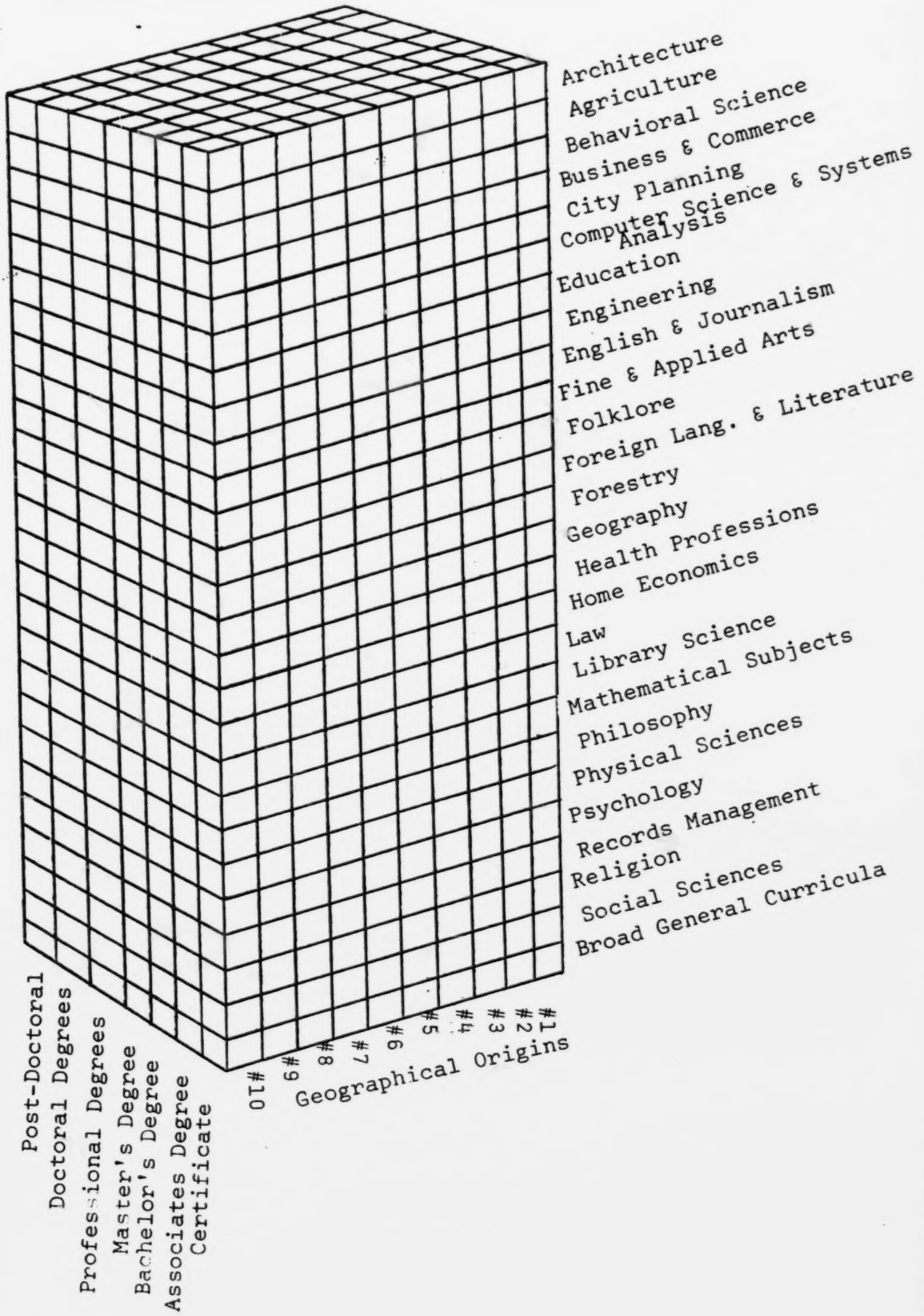


Figure 5

Although the number of the structural components is large, the crux of the problem lies in the availability of current information on the specified categories. This availability, or lack thereof, is at least equal in importance to the relevancy of these classifications. If this amount of detail is necessary for effective planning for higher education in New York, the proposed methodology to be described in Section 3 can be extended to satisfy this objective but its value would be questionable without the necessary data case.

Based on discussions with individuals actively engaged in educational planning within the State Education Department (including the Advisory Committee) and interviews with registrars and planning officials at representative institutions, a <u>feasible</u> input and output morphology was developed and used as a framework for the formulation of the computer model described in Section 3. By reducing the number of categories and consolidating certain items, the total number of data cells has been reduced to a technically feasible quantity. The input now involves approximately 730 cells vs. 4,500 cells originally considered, while the output is 1,640 cells vs. 60,000 cells. From this feasible set of data cells, it is now possible to develop a conceptual model which will consider these data elements.

2.6 The Conceptual Model

The conceptual model will be developed in two phases. First, the micro-structural portion of the educational system will be examined and a micro model developed; then, the macro model will be developed by aggregation.

From the morphological analysis of the inputs and outputs of the total educational system, the inputs and outputs at each level for each curriculum at a particular school can be shown in the form of a conceptual model which appears as Figure 6. [While this is shwon only for a curriculum model, the same reasoning is easily transferable to a college model]. Assuming that the freshman year of a curriculum is represented, the inputs to this particular curriculum are (a) First-time Freshmen, (b) Transfer Students (Internal) - students within the same school who have changed their career goals and are transferring to the particular curriculum, and, (c) Transfer Students (External) - students who enter the freshman class but have attended another college for some time period and thus are not first-time freshmen.

The outputs of this curriculum consist of a student output and an output to the "Outside World". Student Outputs consist of (1) students who progress to the next level of study within the same curriculum, (2) students who do not progress to next level but remain in same curriculum, (3) students who transfer to another curriculum

MICRO STRUCTURE OF CONCEPTUAL MODEL

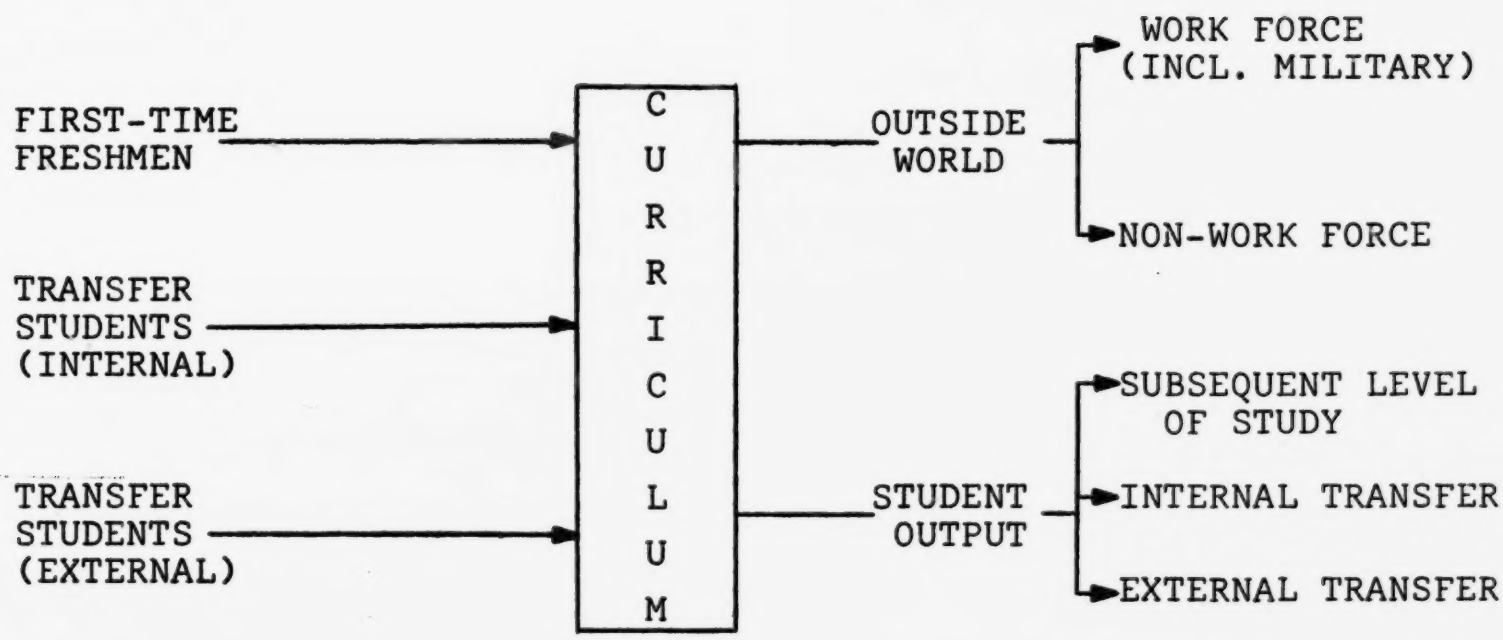
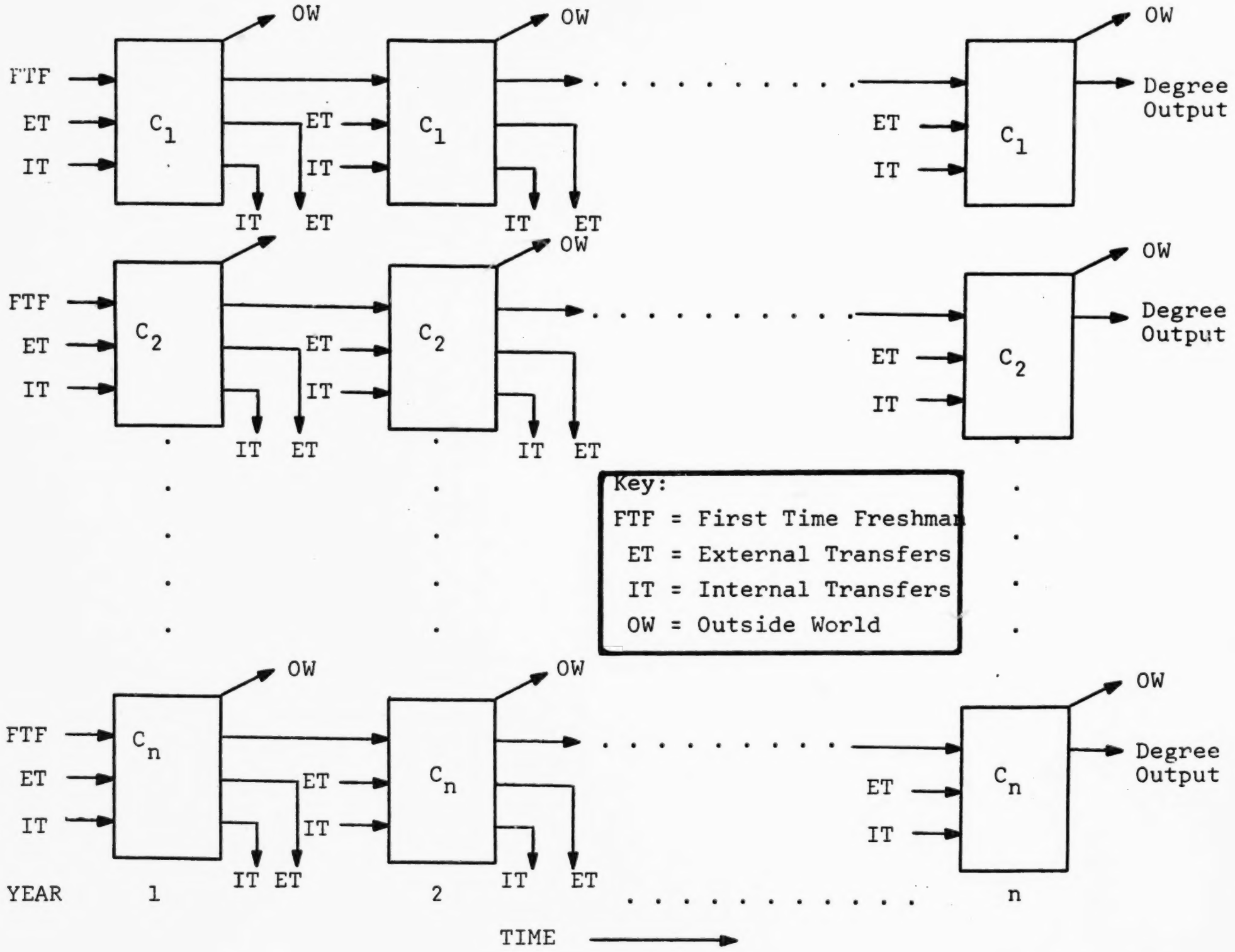


Figure 6

within the same school (internal transfer), (4) students who do not progress to next level and change curricula within same school, and (5) students who transfer to another school (external transfer). Outputs to the "Outside World" are those people who enter the work force or military service and those who enter the non-work force.

The conceptual model for one educational activity within the total educational system can be constructed by aggregating a number of micro-structural models sufficient to describe the total number of curricula within a specified educational activity, i.e., a particular institution of higher education. The resulting model (Figure 7) shows the flow of students through an educational process, by curriculum, and during a specified time span.

It is easy to see that a macro-model of the total educational system could be constructed in a similar fashion by aggregating micro-models of various educational institutions. In fact, if all curricula were aggregated, Figure 7 would represent such a simplified model and would be helpful in analyzing the intercollege flows in the total educational system for purposes of enrollment projection by level of student. The more difficult task of projections by both curriculum and college is also feasible from a theoretical standpoint, but would be extremely difficult to implement due to: (1) magnitude of the data base required, and (2) present computer technology.



THE CONCEPTUAL MODEL

Figure 7

PROPOSED METHODOLOGY FOR PROJECTING STATEWIDE COLLEGE ENROLLMENTS

3.1 Introduction

It is desirable in all simulations to strive for a model which will approach an isomorphism between the model and the environment it is designed to depict. Both the accuracy and relevancy of any simulation are primarily functions of the degree of fidelity achieved. While fidelity is an important factor in the construction of any model, it is of paramount importance in simulation models because of the potential ability of the latter to consider "what if" questions.

Consequently, the model structure which is described in the material to follow accurately reflects many factors now existing in the higher educational environment. The principal theme of the simulation is a representation of the higher educational system by means of a flow model. Students "flow" through the simulated educational system in a manner which corresponds to the real world flows of students in an actual system.

Individual "cells" representing specified characteristics of the students could be treated independently, thus allowing far less sophisticated and more easily computable relationships. However, independent treatment would destroy the homogenity of the model with a corresponding reduction in fidelity, thus negating to a large extent the benefits

In summary, the total educational system has been analyzed from a systems concept, the various segments and classifications which affect the environment of the system have been delineated and a series of conceptual models developed. Using a building-block approach, the basic models can be combined to provide certain aggregate models useful for further experimentation. From this conceptual framework, knowledge of the existing data inputs, and knowledge of the desired informational outputs required for educational planning and decision-making, a computer model capable of simulation can be constructed.

In summary, the total educational system has been analyzed from a systems concept, the various segments and classifications which affect the environment of the system have been delineated and a series of conceptual models developed. Using a building-block approach, the basic models can be combined to provide certain aggregate models useful for further experimentation. From this conceptual framework, knowledge of the existing data inputs, and knowledge of the desired informational outputs required for educational planning and decision-making, a computer model capable of simulation can be constructed.

PROPOSED METHODOLOGY FOR PROJECTING STATEWIDE COLLEGE ENROLLMENTS

3.1 Introduction

It is desirable in all simulations to strive for a model which will approach an isomorphism between the model and the environment it is designed to depict. Both the accuracy and relevancy of any simulation are primarily functions of the degree of fidelity achieved. While fidelity is an important factor in the construction of any model, it is of paramount importance in simulation models because of the potential ability of the latter to consider "what if" questions.

Consequently, the model structure which is described in the material to follow accurately reflects many factors now existing in the higher educational environment. The principal theme of the simulation is a representation of the higher educational system by means of a flow model. Students "flow" through the simulated educational system in a manner which corresponds to the real world flows of students in an actual system.

Individual "cells" representing specified characteristics of the students could be treated independently, thus allowing far less sophisticated and more easily computable relationships. However, independent treatment would destroy the homogenity of the model with a corresponding reduction in fidelity, thus negating to a large extent the benefits

of a simulative approach. By designing a flexible and adaptive model, it will be possible to take full advantage of advances in both computing methodology and evolving data systems. While the proposed methodology is intrinsically a simulation of phases of the New York State higher educational system, its broadest usage will occur when other specialized models (contingency programs) are allowed to interact with this simulation. Thus, the model is useful in itself, but, more importantly, will act as a framework for future analysis.

It will be shown in the following discussion that there are essentially three segments in the model: initialization, operation, and output. Structural aspects will be explained by first providing an overview and then following the operation of the model through means of an abstracted flow diagram [For further clarification a numerical example is provided in Appendix C]. Finally, sections will be presented concerning the projection techniques to be utilized and the concept of contingency programs.

3.2 An Overview of the Model

The purpose of this section is to discuss in a very general way the structure of the proposed model. The methodology can be divided into three separate and distinct segments or phases: initialization, operation, and output.

In the first phase, input information is organized and structured such that projections for future years may be developed through a process of actually "aging" simulated students of previous years. Thus the original input data (in one of several feasible forms as discussed in Section 4.2) is first used for purposes of estimating parameters for projection equations. Once the input data has been ingested by the model, an accurate picture is available of the time sequence of states through which the simulated educational system has passed over a number of years. In this respect, the model is said to be "initialized," that is, it is at this point a static representation of all that has transpired in the educational system for a specified period. (It may be noted parenthetically that the model is self-updating in that as the data for future years becomes available, it is automatically incorporated into the model and, thus, the trend factors are continually updated.) Initialization is, however, not of primary importance since we are interested in the future dynamics of the system — and initialization serves merely to set the stage for these dynamics.

Broadly and briefly, the initialization process gives the model the capability of projecting:

(1) numbers of students at each level (freshman, sophomore, etc.), categorized according to some major classifications scheme such as curriculum or college;

(2) numbers and percentages of students moving from one major classification to another over one-year time periods (for example, the percentage of freshmen at college A in year t who became sophomores at college B in year t+1); and

(3) numbers and percentages of students having particular characteristics within each major classification (characteristics such as age, sex, geographic origin etc.).

Since all of the above quantities generally change over time, they are indexed according to year. In the ensuing discussion, the quantities designated above as (1) are called <u>initial vectors of classification</u>; those designated (2) are combined in ordered arrays to form <u>class-pair transition matrices</u>; and those designated (3) are combined into <u>output breakdown vectors</u>.

It is through the actual operation of the model that the dynamics of the aging process are simulated. For example, multiplication of a year t "initial vector of classifications" for, say, sophomores by a year t "sophomore-junior class-pair transition matrix" yields (conceptually) the initial vector of classifications for year t+1 juniors. (Admittedly, there are more students involved in the initial classification vector for year t+1 juniors, e.g., some students may not have fulfilled their sophomore-level requirements in the educational system under consideration, and may have transferred into our system at the junior

level. The model would, of course, take such groups into account. A series of these multiplications effectively ages each level from freshman through fifth-year student or graduate.

If it is desired to make a finer delineation of the characteristics of the t+1 juniors, each element in the initial vector of classifications for juniors in year t+1 would be multiplied by the appropriate output-breakdown vector for juniors in t+1. This latter operation constitutes the major portion of the ouput segment of the model.

Although it will be discussed at length further on in this paper, it may be efficacious to introduce the concept of the "contingency program" at this point.

As it has been described thus far, the model can answer only the question "if present trends in numbers attending college, transition ratios, and student characteristics continue, what will the college population within the educational system under consideration look like in X years?" Should the user desire to answer questions of the "what if?" category, e.g. "What if draft policies were substantially revised?", the model as proposed, cannot provide a complete answer. Such questions can be answered only by means of subroutines, called by the model's main program, which indicate those projection parameter values to be changed and the magnitudes of the individual changes. These subroutines, called <u>contingency programs</u>,

may be relatively simple to program — but the research necessary to determine which parameters will change and the nature and magnitudes of these changes may prove to be a difficult and expensive task. Thus the model without contingency programs should be visualized as a basic structure whose purpose is to set the dynamics of the simulated system in motion — but whose worth, alone, is to show only future results if present trends continue.

3.3 The Structure of the Model

The following description of the structure of the computer simulation model will use the simplified flow diagram represented in Figure 8 as an outline; the numbers in parentheses will refer to the numbered blocks in the diagram.

Blocks (1)-(26): Initialization of the Model. The computer program based on the proposed model would start by setting the dummy time variable "t" and cause reading by the computer of data on all students who entered the educational system in that first year (t) for which complete data are available, block (2). Blocks (3) through (7) determine the level at which each separate student entered the system, and total numbers of students entering in year t at each level are aggregated and stored for later use. Thus, for example, after all students who entered in year t have been read, EFROSH(t) would give the total number of

freshmen who, although entering our educational system for the first time, are not freshmen per se for the first time (that is, first-time-freshmen or FTF); ESOPH(t) is the total number of students who entered our educational system as sophomores in year t; while EJUN(t), ESEN4(t), and ESEN5(t) are the numbers of juniors, seniors, and fifth-year students who entered our educational system at those levels in year t. Although not apparent in the notation, the above variables are broken down by curriculum or college (depending upon the level of disaggregation of the model), and are thus stored in the form of row vectors with as many elements as there are colleges or curricula being handled by the model. Storage of these variables is indicated by blocks (8)-(12). In addition, FTF(t) (first-time freshmen who entered the system in year (t) are stored in a similar form [block (13)].

In blocks (14)-(17), it is noted that the value of t is decremented by 1, 2, 3, or 4, depending upon whether the student entered the system as a sophomore, junior, senior, or fifth-year student, respectively. While this step is not essential for the operation of the model, programming and compiling time can be greatly decreased if the same two routines can be used for development of transition matrices and output-breakdown vectors for all student levels. The nature of the model — the fact that it is a flow-model — requires that students whose data contribute to the development of a year t "freshman-sophomore

transition matrix" must also have their data contribute to the development of a year t+1 sophomore-junior transition matrix, a year t+2 junior-senior transition matrix, and so on, unless they (particular students) leave the system in the interim. If one transition-matrix development routine is written, then all (say) sophomore-junior transition matrices will involve the year t+1 and t+2, if the students being analyzed started as freshmen (i.e., the students are sophomores in t+1, juniors in t+2). Obviously, the student who starts as a sophomore in year t cannot be used in the development of a year t+1 sophomore-junior transition matrix, but must be used in a <u>year t</u> sophomore-junior transition matrix. Thus we decrement the value of t by one when data from a student who started as a sophomore are used in transition-matrix development. It may be noted at this time that it is not impossible, according to the proposed model, that a sophomore entering the system in year t be used in developing a t+1 sophomore-junior transition matrix: the latter would merely indicate that the particular student somehow did not fulfill the requirements for the successful completion of his initial sophomore year in our educational system — and as a result, he was again a sophomore in the following year. As will be seen, such student difficulties are left out of the flow-chart (although they will be incorporated into the model) for the sake of clarity.

In block (18), the transition matrices are developed for all class-pairs. For future reference, FS(t), SJ(t), JS4(t), and S4S5(t) are defined as transition matrices for the class-pairs "freshmen-sophomore," "sophomore-junior", "junior-senior", and senior-fifth-year", respectively, the t subscript indicating the chronological association of each transition matrix. As the model is presently visualized, if it operates on a curriculum basis, input information will be in the form of unit records; if, on the other hand, it operates on a college or campus basis, input will be one order of aggregation higher — that is we would input "total numbers of students transferring from college A to college B as (say) sophomores in year t." These numbers would, however, be handled in the exact same way by the program. Since a row in the matrix shows the relationship between "where a particular class of campus A students were at time t, and where at time t+1" the raw row total would be the total number of campus A students at time t, and each element would show how many of this total were at campuses A, B, C, etc. at time t+1. Dividing each element by the raw row total (of the row in which that element is found) gives these flow relationships in percentage terms. In essence, development of initializing transition matrices is merely repetition of the latter process for all levels and years. Since it is expected that beginning usages of the model will involve data in percentage terms, the latter may be injected directly into the appropriate positions in

the transition matrices, although this is not shown by the flow diagram.

In block (19), approximately the same process is used for development of output breakdown vectors as was used for development of the transition matrices. Again, however, while the transition matrices give some indication as to the flows through the system, the output-breakdown vectors' sole purpose is to indicate the characteristics of certain classifications of students (whether classified by curriculum or college.) For future reference, FROBV(t), SOOBV(t), JUOBV(t), S4OBV(t), and S5OBV(t) represent Output-Breakdown Vectors of freshmen, sophomores, juniors, seniors, and fifth-year students respectively in year t. Although the notation does not indicate the fact, each of these vectors involves only one level but as many classifications of students as the model will be handling. Thus if a ten college model is used, FROBV(t) will actually involve ten output breakdown vectors - one for freshmen at each of the schools.

In block (20) the model checks to see whether all data have been read. If such is not the case, t is incremented by 1 [block (21)] and data are now read for the subsequent year. If, on the other hand, all data have been read, the program proceeds to block (22) which begins determination of the parameters to be used in our future projections. First in line is the determination of projection parameters

for transition matrices. Having read a number of years' worth of data, a like number of freshmen-sophomore, sophomore-junior, etc., transition matrices — one for each year t, t+1, t+2, etc., is available. Thus, for example, if five years' worth of data are initially read, there would be five freshmen-sophomore transition matrices [that is, FS(t), FS(t+1), FS(t+2), FS(t+3), and FS(t+4)], five sophomore-junior transition matrices, etc. While the class-pair matrices would not differ in size, the value of any given element within one set of the matrices [say, the set for sophomore-junior transitions, i.e., SJ(t+1), SJ(t+2), SJ(t+3), SJ(t+4), and SJ(t+5)] could conceivably change over time, and it is this change which must be projected into the future. If, for example, a sophomore-junior transition matrix had two-hundred elements in it, five years' worth of data would give five data points for each element — a total of one-thousand pieces of information concerning sophomore-junior transitions alone. Each group of five would be the basis for projections as to the future values of each separate transition element. Naturally, the same method would be followed in the projection of elemental values of the sophomore-junior, junior-senior, etc. transition matrices. While actual future transition matrices would not be developed at this point in the program, the capability of projecting them would be implicit in the fact that the parameters for each separate projection

would at this point be developed and stored. In the same manner, projection parameters would be developed (block 23) for projection of all output breakdown vectors. Again if five years' worth of data were available, each element in each output breakdown vector would have a group of five data points associated with it from which projections as to future values of the elements could be inferred.

As will be recalled, blocks (8)-(13) aggregated the quantities FTF(t), EFROSH(t), ESOPH(t), EJUN(t), ESEN4(t), and ESEN5(t), that is, the number of first-time-freshmen, number of entering freshmen who were freshmen elsewhere, number of students entering as sophomores, juniors, seniors, and/or fifth-year students in year t. In block (24) the program determines projection parameters for these quantities. Again, with five years' worth of data, "t" will have moved from 1 to 5; therefore, for each of the above vectors, five data points would be available for the projection of each element within them.

In block (25) DEGRE4(t), and DEGRE5(t) are vectors of percentages of students in particular classifications receiving degrees upon their leaving school. Essentially, these percentages could be determined as a special part cf the output breakdown vectors and would be indexed by year.

Block (26) is more fully covered in the section concerning contingency programs and their conception. We

note that it would be at this point in the model that any available contingency programs would be called such that they could operate on the projection parameters as desired.

Blocks 1-26 complete the initialization of the model. Since we are working with a simulation model which describes quite accurately the actual ongoing functions of an educational system, the ability to project the operation of a year t+5 requires that we know what went on in the years t+4, t+3, t+2, and t+1. More specifically, to project certain desired data concerning seniors in year t+4, we must certainly age t+1 freshmen through the model for it is basically they who are the t+4 seniors. Given the five years of data, t through t+4, that have gone into the computer for the development of projection parameters, we have, in effect, an ongoing model of the educational system under consideration and thus output concerning year t+5 can be given. On the other hand, output concerning <u>all levels</u> at year t+6 cannot be given since we have not yet aged any of the year t+5 freshmen. The point is, at any rate, that since we <u>do</u> have the first five years' worth of data, the model has been "initialized" and the aging process can begin.

<u>Blocks (27)-(52): Operation of the Model</u>. The purpose of this model does not begin to be realized until projections are actually made with it. Thus the next direct step after block (27) is the routine concomitant to the

aging process itself. The aforementioned block is merely a dummy for the purposes of explanation: we find that projections of a cross-sectional view of students in the educational system in year X are desired.

Since a cross-sectional view of students for year X is desired, we must begin the flow process with those students who will be fifth-year students in that year, that is, the freshmen of year X-4. While the word "calculate" is used in block (28) with regard to determination of year X-4 data, it is possible that X is so close to the years from which the (hard) initialization data came from, that the values of FTF (X-4) and EFROSH(X-4) needn't be projected, but can read directly out of core or from tape. If, on the other hand, year X is so far in the future that no hard data exist on X-4 freshmen, then values must be projected for them. Block (29) tells us that the total number of freshmen (again classified by college or curriculum) in year X-4, that is, TFROSH(X-4) is simply equal to the sum of the first time freshmen in year X-4, the number of freshmen coming from another educational system to take their freshman year over again, and internal transfers to the freshman level. From this point, it will be assumed that year X is so far in the future as to force calculation of data for year X-4 rather than offering the opportunity of reading it from hard data.

In block (30), the program calculates the freshman-sophomore transition matrix for year X-4. Strictly speaking, there are two years associated with each transition matrix, and in this case the years would be X-4 and X-3. The foregoing simply means that an X-4 freshman making a transition to some classification at the sophomore level will be at the sophomore level in year X-3. In block (31), the user is given the option of printing out the varied characteristics of the freshman class of year X-4 — that is, how many in age group 1; how many male; etc. In the following block (32) the year X-4 freshman-sophomore transition matrix [FS(X-4)] is multiplied by TFROSH(X-4), and if it will be recalled that TFROSH represents a vector of values, it is seen that the result of this multiplication will itself be a vector. This vector, called SOPH with appropriately indexed year (in this case, X-3) is by far the greater proportion of the total number of sophomores in year X-3 within our educational system. The remainder of the sophomores are obtained from the projected value of ESOPH(X-3) (block 33) — the number of sophomores projected to enter the system as sophomores in the year X-3. Blocks (34)-(37) perform the same function and calculations as were carried out in blocks (29)-(32). In the present case, however, a single level higher is being analyzed, and the indexed year must follow suit. Thus freshmen in year X-4 were aged into sophomores for year X-3, and multiplying the vector of total

sophomores (TSOPH(X-3)) by the sophomore junior-transition matrix (SJ(X-3)) the result is the major part of the juniors of year X-2, block (37). Again [in block (38)] the number of year X-2 juniors who entered as juniors must be calculated and added to JUN(X-2) block (39), to give the total junior class (by appropriate classification) for year X-2, that is, TJUN(X-2). Although not explicitely mentioned for the case of sophomores, if the user desires, the characteristics of this junior class can be given by multiplying each element of TJUN(X-2) by the appropriate output breakdown vector for year X-2, block (41), in this case JUOBV(X-2) — which, it is recalled, has as many vectors associated with it as there are relevant classifications (curriculum/college.)

Blocks (42) through (49) perform the same calculations for higher levels that were performed for the freshman and sophomore levels. Since we are, in essence, aging the same group of individuals with minor additions from outside the educational system, as they progress to higher and higher levels, the year index must increase in value. By block (49), the freshmen class of X-4 has been aged to fifth-year students (although it should be noted that in block (49), many of the X-4 freshmen were lost as they graduated from a four-year program and entered the outside world). The percentage of these fifth-year students obtaining degrees is determined, block (51), and this class's various characteristics are printed out after multiplication of the

vector TSEN5(X) by the appropriate output breakdown vectors S50BV(X) - block (51).

Blocks (53)-(55): Output of the Model. Recall that the original desire was a cross-sectional representation of the students in year X; at this point in the program, only the fifth-year students of year X are available. Thus a freshman class from year X-3 must be aged to year X; and the result will be year X's senior (fourth-year) class.. Aging a freshman class from year X-2 to year X will yield year X's junior class, while year X-1 freshmen will yield X's sophomores. Finally, projection of FTF(X) and EFROSH(X) will give a view of year X's freshman class — and the desired cross-section has at last been developed. Along these lines then, the program (at block 53) asks whether the first set of projections was for year X. At this point in the explanation, only fifth-year students for year X have been calculated since the first set of projections for freshmen was for year X-4. As a result, the model updates the value of X by 1 and returns to block (27) whereupon, in effect, the freshmen from year X-3 begin to be aged.

Although not shown on the flow chart, either a storage subroutine or a subroutine to check the value of X for the freshman projections is necessary to (in the case of the former) store the extra data generated by the model or (as in the case of the latter) stop the freshman aging process at each point when the desired output year is reached. The

fact that X was updated by 1 in block (58) indicates implicitly that the same routine will be used for the development of the necessary projections. As a result, when the model begins to age the freshmen, it will age them until they are fifth-year students. If a cross-section at year X is desired, freshmen which the model begins to age in X-2 will be output as juniors — however, the model will continue to age them until they become fifth-year students — and it is these data which may be considered extra. The easiest method would be to store these extra data — since if at some time the user desires projections for year X-1, most of the work is already done; all that remains is to project the freshman class for year X+1, and read the stored data from calculations of student cross sections for the previous year.

In block (55), the model asks whether another year's cross section is desired by the user. Having stored the data, such an output would be a relatively simple matter to produce as was shown above. If no other cross sections are wanted, the program ends. The reader should note that although only a cross-sectional output was discussed, the varieties of output format are nearly limitless.

The foregoing discussion is of necessity technical in its nature. However, an understanding of the structure of the model is necessary for an effective evaluation of its usefulness to higher educational planners. Appendix B

presents a simplified example for additional clarification.

3.4 The Projection Methodology

The proposed methodology will employ weighted projection techniques throughout. Naturally, the form of any particular projection will be a function of the data involved. Each parametric function projected will be such that it provides "the best fit" to the data used, and consequently each projection is to some extent unique. However, it will be necessary to incorporate one particular mode of projecting throughout the model in order to satisfy programming constraints.

There is by no means unanimity as to the "best" projection technique to use in an educational environment. Witness, for example, the amalgam of techniques discussed by Lins (L. J. Lins, Methodology of Enrollment Projections for Colleges and Universities: A.A.C.R.A.O., 1960). The most often utilized technique is one which incorporates a particular age group (say, 18-21) as a base and then projects some college characteristic as a direct function of the size of this age group. While it is true that a large percentage of college students are between 19 and 21 years of age, Lins states that the percentage of college students between these limits rarely exceeds 70%. Moreover, while projections which use age group as a base may prove remarkably accurate in the aggregate, these projections

FLOW CHART OF PROPOSED COMPUTER MODEL

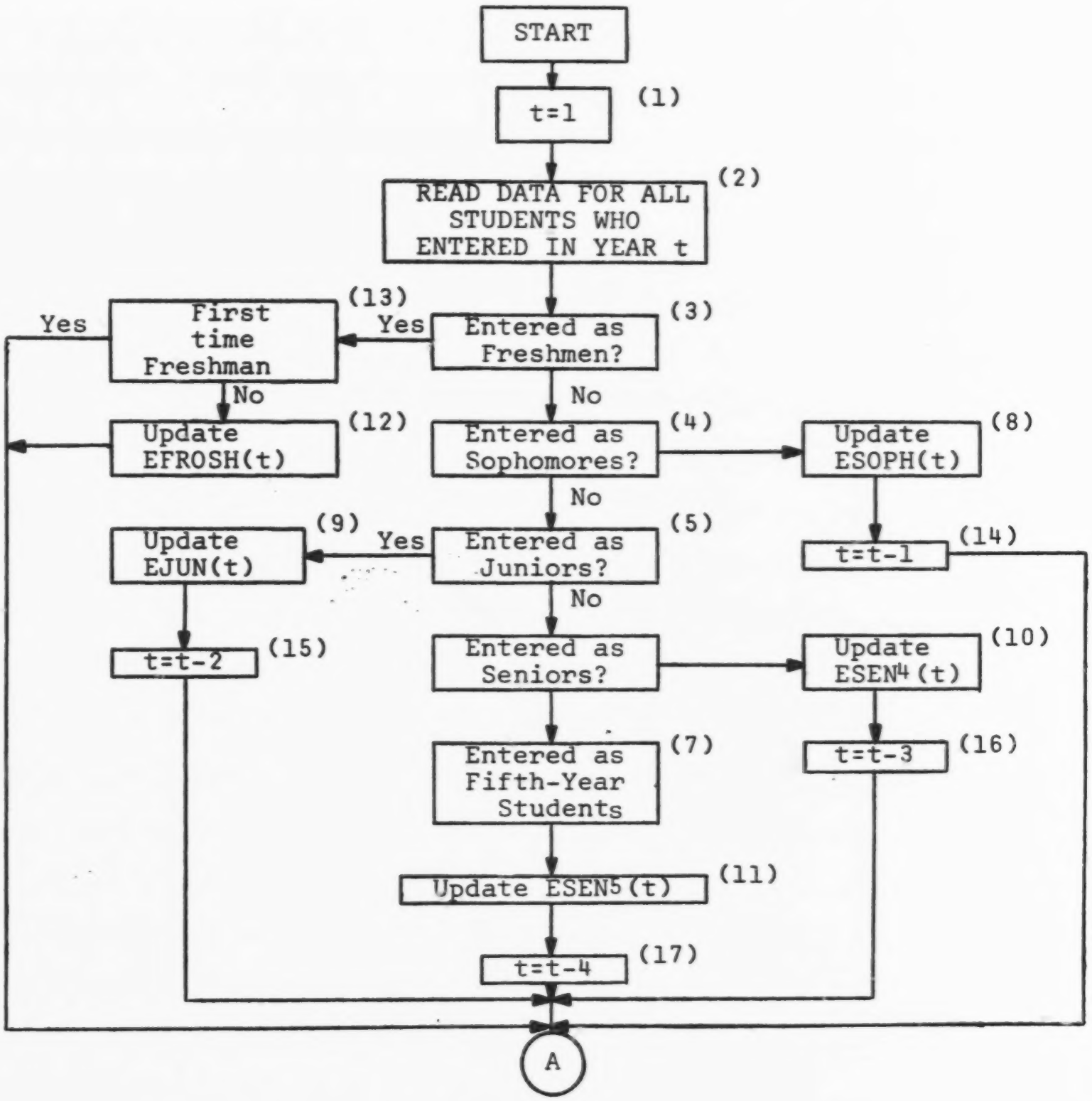
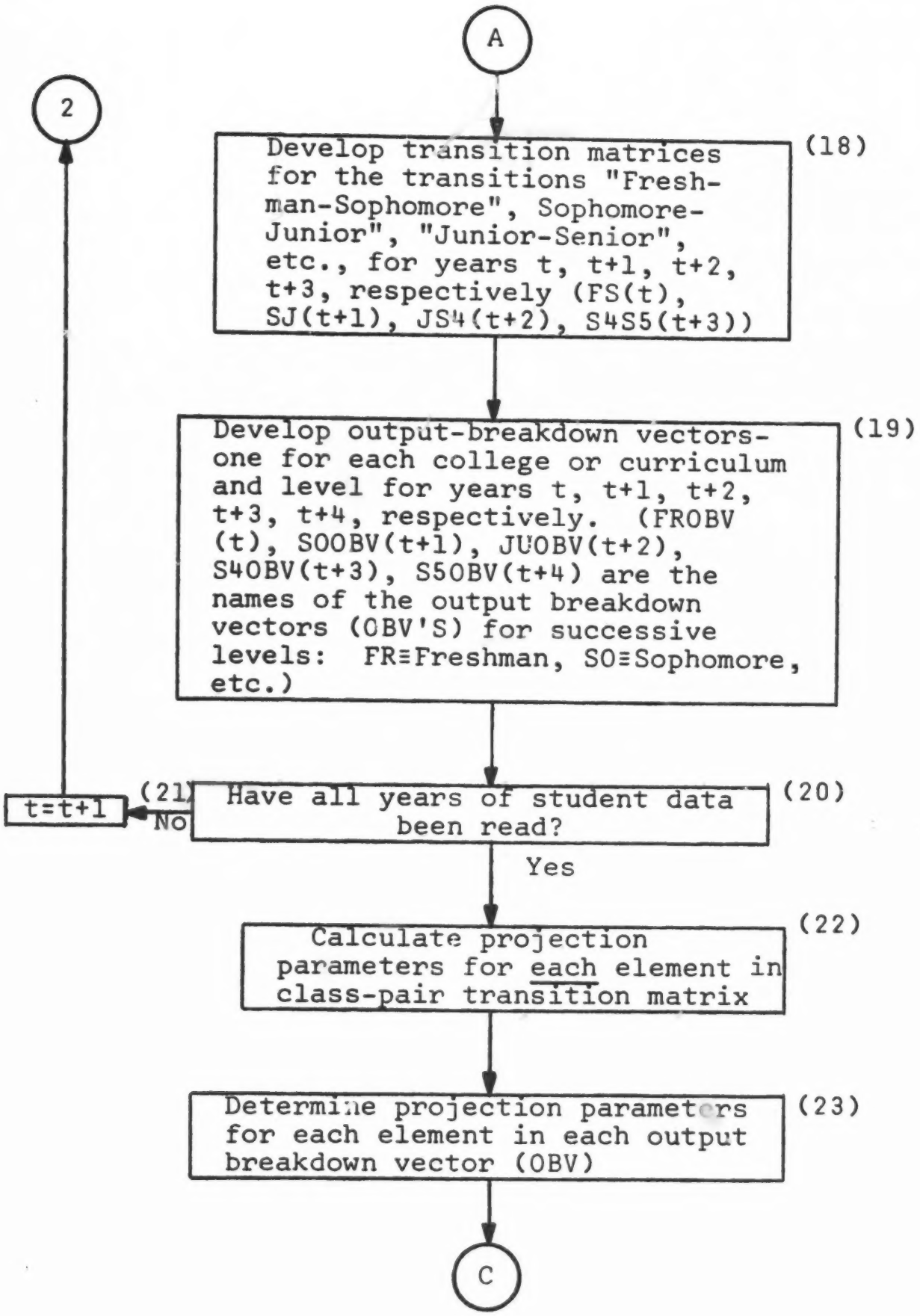
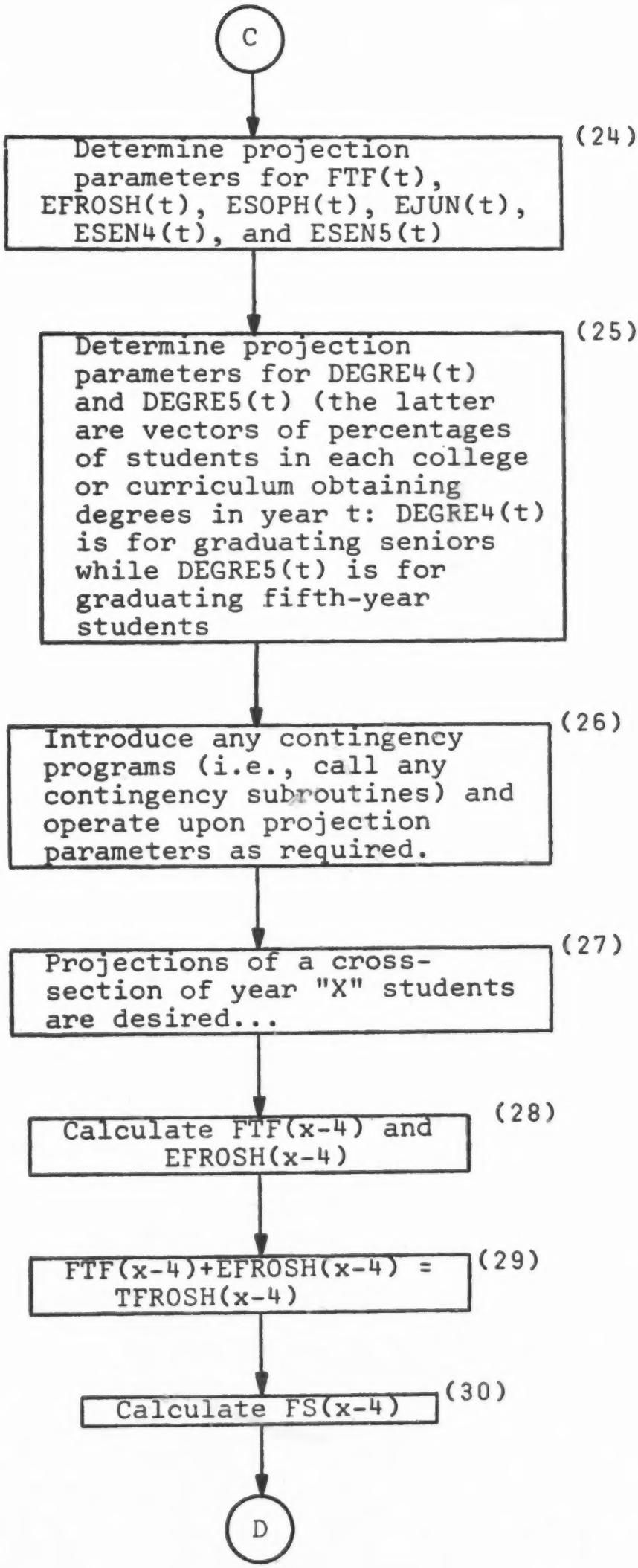
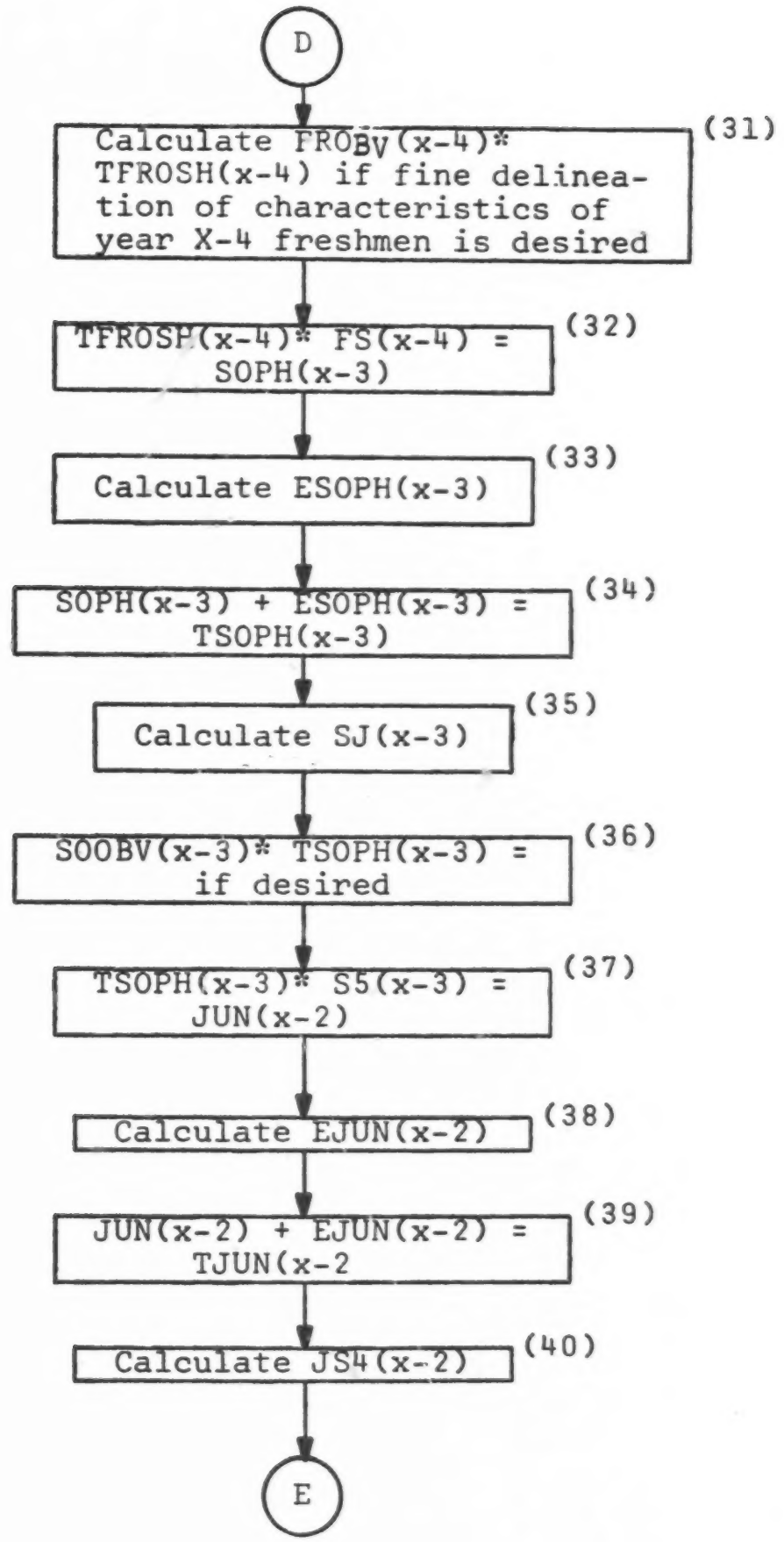
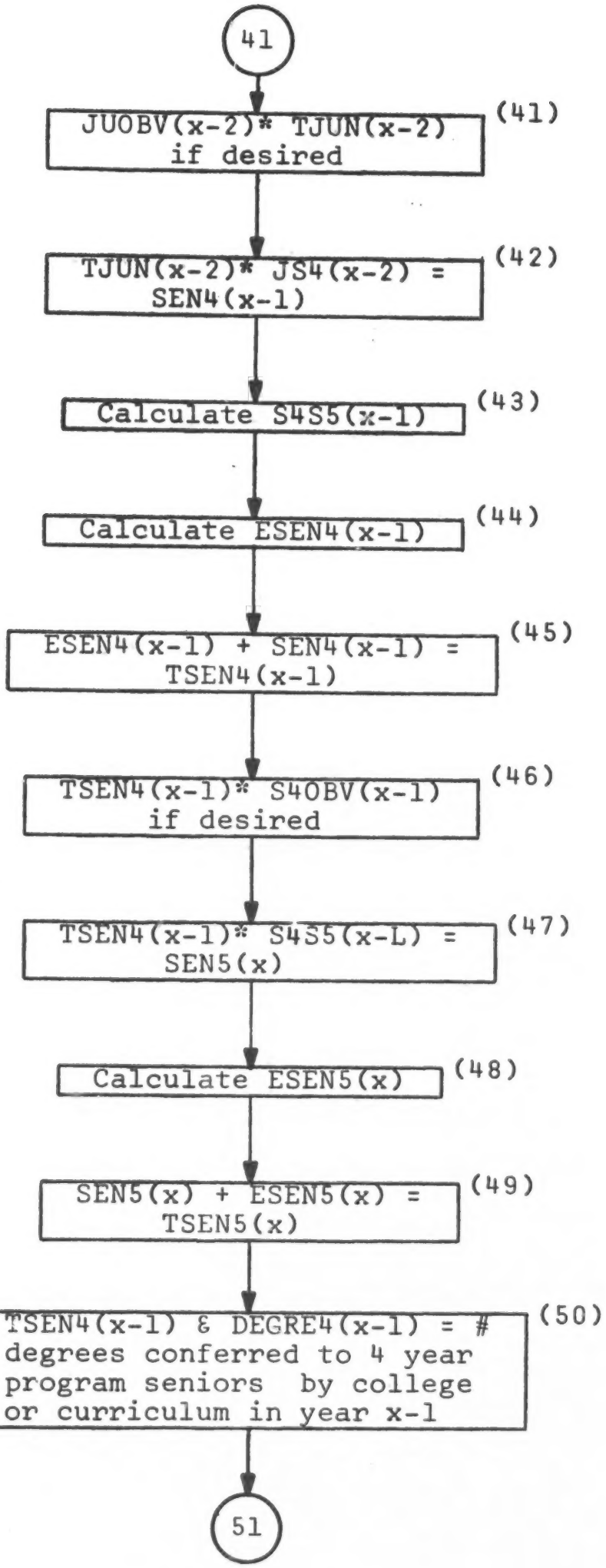


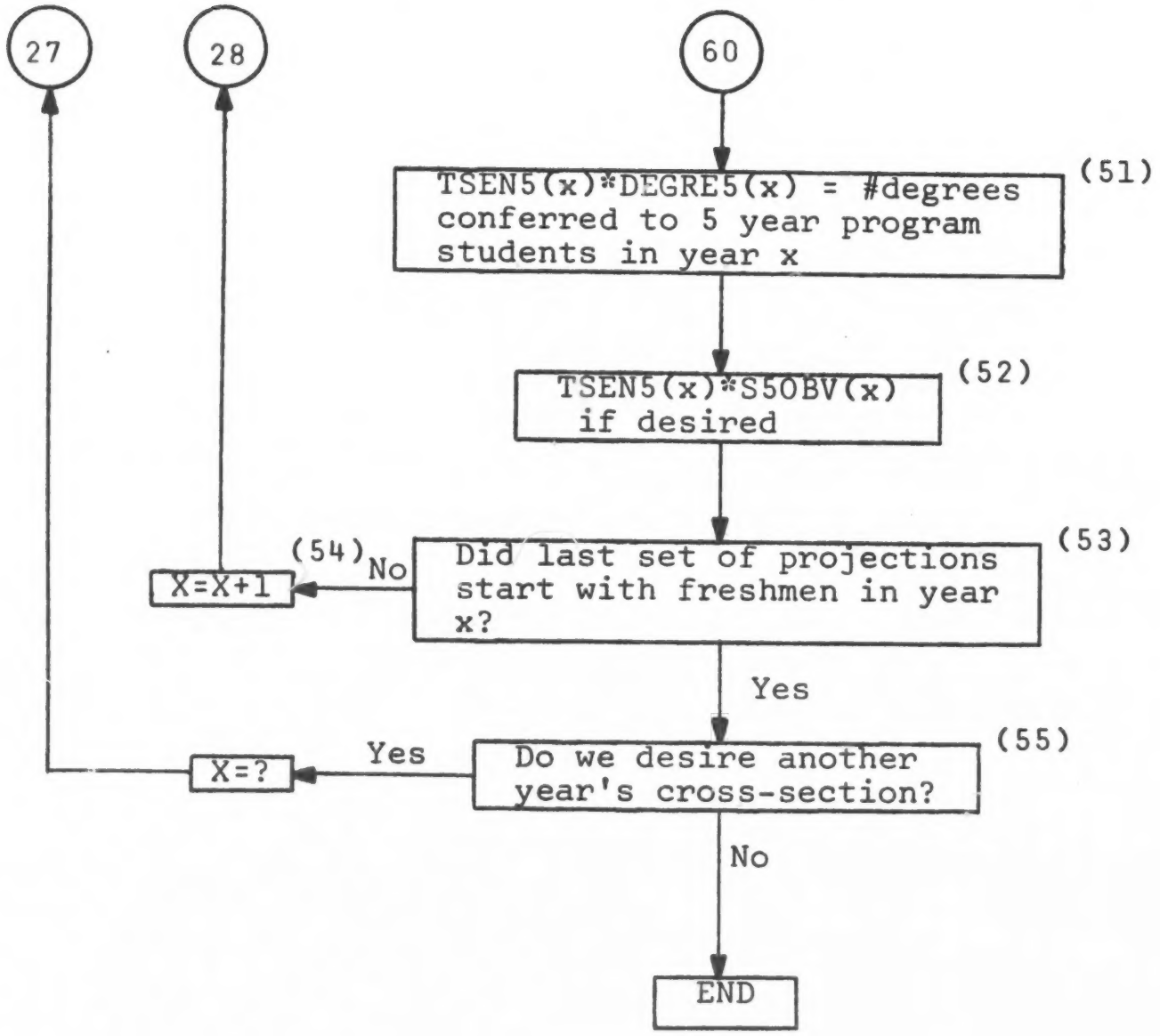
Figure 8


A
2
Develop transition matrices for the transitions "Freshman-Sophomore", Sophomore-Junior", "Junior-Senior", etc., for years t, t+1, t+2, t+3, respectively (FS(t), SJ(t+1), JS4(t+2), S4S5(t+3))
(18)
Develop output-breakdown vectors-one for each college or curriculum and level for years t, t+1, t+2, t+3, t+4, respectively. (FROBV (t), SOOBV(t+1), JUOBV(t+2), S4OBV(t+3), S5OBV(t+4) are the names of the output breakdown vectors (OBV'S) for successive levels: FR≡Freshman, SO≡Sophomore, etc.)
(19)
Have all years of student data been read?
(20)
No
(21)
t=t+1
Yes
Calculate projection parameters for each element in class-pair transition matrix
(22)
Determine projection parameters for each element in each output breakdown vector (OBV)
(23)
C


C
Determine projection parameters for FTF(t), EFROSH(t), ESOPH(t), EJUN(t), ESEN4(t), and ESEN5(t)
(24)
Determine projection parameters for DEGRE4(t) and DEGRE5(t) (the latter are vectors of percentages of students in each college or curriculum obtaining degrees in year t: DEGRE4(t) is for graduating seniors while DEGRE5(t) is for graduating fifth-year students
(25)
Introduce any contingency programs (i.e., call any contingency subroutines) and operate upon projection parameters as required.
(26)
Projections of a cross-section of year "X" students are desired...
(27)
Calculate FTF(x-4) and EFROSH(x-4)
(28)
FTF(x-4)+EFROSH(x-4) = TFROSH(x-4)
(29)
Calculate FS(x-4)
(30)
D


D
Calculate FROBV(x-4)* TFROSH(x-4) if fine delineation of characteristics of year X-4 freshmen is desired
(31)
TFROSH(x-4)* FS(x-4) = SOPH(x-3)
(32)
Calculate ESOPH(x-3)
(33)
SOPH(x-3) + ESOPH(x-3) = TSOPH(x-3)
(34)
Calculate SJ(x-3)
(35)
SOOBV(x-3)* TSOPH(x-3) = if desired
(36)
TSOPH(x-3)* S5(x-3) = JUN(x-2)
(37)
Calculate EJUN(x-2)
(38)
JUN(x-2) + EJUN(x-2) = TJUN(x-2
(39)
Calculate JS4(x-2)
(40)
E


41
JUOBV(x-2)* TJUN(x-2) if desired
(41)
TJUN(x-2)* JS4(x-2) = SEN4(x-1)
(42)
Calculate S4S5(x-1)
(43)
Calculate ESEN4(x-1)
(44)
ESEN4(x-1) + SEN4(x-1) = TSEN4(x-1)
(45)
TSEN4(x-1)* S4OBV(x-1) if desired
(46)
TSEN4(x-1)* S4S5(x-L) = SEN5(x)
(47)
Calculate ESEN5(x)
(48)
SEN5(x) + ESEN5(x) = TSEN5(x)
(49)
TSEN4(x-1) & DEGRE4(x-1) = # degrees conferred to 4 year program seniors by college or curriculum in year x-1
(50)
51


27
28
60
TSEN5(x)*DEGRE5(x) = #degrees conferred to 5 year program students in year x
(51)
TSEN5(x)*S50BV(x) if desired
(52)
Did last set of projections start with freshmen in year x?
(53)
(54) No
X=X+1
Yes
Do we desire another year's cross-section?
(55)
Yes
X=?
No
END

become increasingly poor as we disaggregate; mainly due to the almost complete lack of functional connectiveness of this technique and the college populations.

Of course many other techniques are also currently available, such as cohort-survival, ratio method, curve fitting method, correlation analysis and exponential smoothing. A close inspection of these methods reveals that they all may be incorporated under the general heading of weighted regression analysis. Naturally, some special formats are necessary to incorporate each of these possible methods in a particular regression program.

There are three basic reasons for using weighted regression analysis in the construction of the computer model under consideration. First, weighted regression analysis allows flexibility in the functional forms, thus permitting the accuracy and precision of the estimates to improve automatically as better data sources become available. Consequently, while certain sets of projections are likely to have the same form, the individual projections within sets will differ. Secondly, this approach permits the systematic adaptation of the model to possible changes which might be envisioned, as described in Section 3.5. This means an entire set of projections can be varied, since they share common forms, while still allowing differences in the effect of contingency programs upon other projection sets. Finally, the projection mechanisms

employed must fit within reasonable computability constraints. A great deal of information concerning the computational characteristics of various weighted regression routines is available.

3.5 The Concept of Contingency Programs

The notion of "contingency programs" has occurred several times throughout this section of the report. Although the concept is to some extent self-explanatory it behooves us to clarify the term before examining possible ramifications. Contingency programs are separate computer models which will systematically (either stochastically or deterministically) change parameters in tne student flow model in order that the effect of various changes in the environment may be reflected in its output. Potentially each contingency program will act upon parameters of the student flow model in a specific manner.

For instance, it may be desired to note the changes in flows of students which might be expected because of a change in an environmental condition, e.g., a major economic recession. An analysis would be undertaken of the possible effects such a change might have upon the various flows within the model and a program would be constructed to act as an operator upon the relationships in the model. The output of the model — incorporating this contingency program — would then provide an estimate of the overall

effect the recession might have on higher education in New York State. A separate computer program could be constructed to handle each particular question which might arise.

While contingency programs are of obvious value, the construction of such programs may prove to be difficult. The primary reason for the difficulty of constructing a contingency program resides in the nature of the "what if" question. It is unlikely that there will be sufficient data readily available upon which conclusions can be drawn as to the effect of hypothetical environmental changes. The impact of these environmental changes might, for example, be assessed by knowledgeable individuals and this information used as a basis for constructing the appropriate sub-model.

The use of contingency programs will have the additional benefit of requiring a clear statement of the problem and a careful assessment of possible ramifications of proposed environmental changes by the user of the simulation model.

The proposed methodology is such that future contingency programs will be easily incorporated into the present structure. As an example, let us consider the elements of the model which might be affected by a particular contingency program, e.g. one constructed to assist in evaluating a proposal to decrease financial aid to the science and engineering fields and increase aid to liberal arts fields. A careful analysis of the effects of

such a proposal on the disaggregated segments of the model could indicate that a shift toward arts curricula and away from science curricula might be expected. Furthermore, one might expect that this effect would be noticed among incoming freshman, t years hence, and that this trend would percolate through the entire system. A contingency program could then be written which would reflect these expectations in altered student flows. A run of the model incorporating this contingency program would provide estimates of the ultimate effect of this possible shift in support on all segments of the final output.

Although this concept is presently at the formative stage, it should be possible to construct various levels of contingency programs. Since the specificity and exactness of a contingency program are a direct function of the effort (translatable into dollars and time) expended in its construction, various programs — each at a particular cost — could be considered and a cost/benefit analysis, such as that described in Section 4.3, could be performed. The nature of a simulation makes it possible to evaluate the effects of a wide range of environmental changes with the same contingency program. This approach then permits a sensitivity analysis through systematic variation of parameter values within the contingency program.

3.6 Summary

This section has presented the proposed methodology for projecting statewide college enrollments. Starting with a discussion of the advantages of the simulation approach, an overview and a detailed, technical presentation of the structure of the model were given. The proposed simulation model uses the concept of student flows and describes these flows by means of initial classification vectors, transition matrices, and output breakdown vectors.

Since it is necessary to estimate the relevant parameters in the model, a projection methodology, employing weighted regression analysis, was suggested. Finally, the concept of contingency programs was described. This characteristic of the proposed methodology permits the assessment of "what if" questions, thus utilizing the prime advantage of the computer simulation approach.

FEASIBILITY OF PROPOSED METHODOLOGY

The purpose of this section is to present a feasibility analysis of the proposed methodology for projecting statewide college enrollments. A discussion of the technical feasibility is followed by an analysis of the information system needed to support the proposed computer model. A cost/benefit relationship is also presented that provides a conceptual framework for detailed analysis of the present structure of the model as well as future configurations.

4.1 Technical Feasibility

As outlined in the previous section, the computerized model for projecting statewide college enrollments is technically feasible. Using the proposed methodology, a simulation model should be programmed and can be operated on any computer system having sufficient memory capacity; the constraints would be the level of disaggregation required and expenditures for computer operations. For example, the model described in the previous section with ten to twenty basic classifications (college or curricula) and providing up to eighty possible characteristics of the students including (five levels of) age, sex, 65 geographical origins - etc., is programmable without use of tape on an IBM 360/50 system with a total core memory of 1.28 million bytes.

An important feature of the simulation approach is the capability of introducing contingency programs that would be designed to show the potential effect of environmental change on the educational process. Quantitative answers to hypothetical questions such as "what would be the effect on future enrollments of limitation of selective service deferments to only science and engineering majors?" can be provided to educational planners. The proposed methodology permits the addition of a subroutine or "contingency program" with very minor reprogramming of the basic model.

One of the most important advances in computer technology has been the development of "on-line" and "real-time" capabilities, where "on-line" refers to the ability to enter input data into and receive transmitted data from the computer directly from the point of origination, and "real-time" means that data can enter, be processed and transmitted by the computer fast enough to affect the functioning of an external environment at that time. For example, a committee studying construction plans for a new campus might want a series of enrollment projections based on different environmental conditions. An on-line, real-time capability would enable the group to evaluate the effects of their different assumptions during a much shorter period and, therefore, be able to make a decision much sooner. The proposed methodology does not have any unique characteristics that would prevent its being used in an on-line, real-time computer system.

4.2 The Information System: A Critical Factor

One result of this study was the realization that there is another very important aspect to the problem of projecting enrollments: that of an accurate and timely flow of information on the status and composition of the college-going population in the State. Without current data characteristics of this population such as those listed in Section I of this report (for example, geographic origin of first-time college freshman), the reliability of any model designed to project such information is higher questionable on a statewide basis.

It is possible, however, to concurrently develop the projection model and an information system, and, in fact, it is extremely desirable that these activities be closely coordinated. For example, in the process of constructing a pilot model at some representative institution, the form and content of existing primary data at that school could be determined and possible methods of collection and summarization can be tested. In addition, the simulation model was designed to permit evaluation of the sensitivity of its output to the input information and estimated parameters. This could aid the designers of an information system in determining the required degree of recency and level of accuracy of the data. Also, the definition of the existing educational environment as discussed in Section I will serve as a starting point for a similar determination

in the development of an inclusive information system.

As previously stated, the objective of this study was to propose a methodology that would result in a model designed not only for today's data but which would grow and improve with the information system. This is possible because of the incorporation into the model of weighted regression equations whose parameters can be modified as the data base is changed. Also, the ability to add subroutines permits adapting the model to different input formats.

Within the structure of the model, there are three basic data requirements: (1) estimation of the parameters within the model; (2) projection of input data; and (3) estimation of output characteristic parameters. Based on our interviews with registrars and planning officials from a sample of colleges in New York State (see Appendix A), these requirements can be satisfied at three levels: (1) asking for subjective estimates from knowledgeable people; (2) statistically sampling records at representative institutions; and (3) developing a higher education data bank including information on student flows.

The first proposed method, obtaining subjective estimates, is definitely economically feasible. This approach would involve interviewing registrars and asking questions such as "What percent of your students are repeaters at the freshman level?" or "What percent of your students come from Nassau County?" Answers might take

the following form: "about 2%" or "a little over 10%". This type of information can then be used to develop initial estimates; the simulation model can be run and the output evaluated for its sensitivity to changes in these estimates. In addition, these data are extremely helpful (and may be necessary) in designing statistical sampling plans.

The next level of data system suggests the designing of sampling plans permitting statistical evaluation of the results. These plans (an illustration of which is discussed in Appendix C) would require careful delineation of the information requirements and the form and content of the data base. It is technically feasible to develop such a plan but the method to be employed (and effort required) depends on the form of the information. Some institutions have student records in a readily accessible form such as tape or punched cards; others have only the unit student record or admissions form. Initial evaluation at Rensselaer showed that it was advisable to obtain subjective estimates of the magnitude of the values of such parameters as inter-curricular student flows prior to the development of sampling plans. It is possible, for example, to obtain estimates of student flow among colleges in New York State by: (1) stratifying the colleges into (say) private-four year, CUNY, SUNY, Private-two-year, Agricultural and Technical colleges, Community colleges, and others; (2)

asking planning officials at the State Education Department, State University of New York, City University of New York, and other representative institutions for their subjective estimates of the flows; and (3) designing a stratified sampling plan to obtain the necessary statistical estimates. As a result of this sampling, weighted regression equations of the form discussed in Section 3.4 could be developed and sampling conducted at regular intervals to check the accuracy and precision of the estimates and make revisions where necessary.

At the highest level of sophistication, a higher education data bank, the proposed computer model will make its greatest contribution. The adaptive properties of the weighted regression equations will permit greater accuracy and precision in estimating parameters; comparison with actual data can be made for evaluation and revision, if necessary, of the structural equations; and finer classification of the output information will be possible.

It was not within the scope of this study to recommend the nature of this total information system. The methodology proposed does not depend for its usefulness upon a particular type of information system (or a specific computer configuration) but does increase in usefulness as the input data improve. However, the knowledge gained of the characteristics of the higher educational environment in New York State through construction of the simulation model

should prove extremely useful in designing the related information system.

4.3 A Suggested Cost/Benefit Relationship

The utility of any system must be determined by relating its cost to the associated benefits. Ideally, all aspects of the utility should be expressable in quantifiable form. When dealing with planning in any socio-economic environment, however, this is seldom possible. Nevertheless, an attempt should be made to interrelate analytically those aspects of the problem that are quantifiable, and delineate for decision-makers those portions which require judgmental evaluation.

The construction of a computer simulation model is similar to acquiring a piece of capital equipment; there are large fixed charges that must be distributed over its useful life. In the case of a computer model to project college enrollments, the fixed cost of construction depends on the following factors: (1) the level of disaggregation desired by the users; (2) the type, nature, and number of the contingency programs; and (3) the form of the input data.

Of the foregoing factors, the most important is the level of disaggregation. The costs of constructing a model to project only first-time freshman on a statewide basis vs. one capable of projecting a desired output such

as "number of economically disadvantaged, female, freshman, biology majors of age 18 from Nassau County at SUNY/Buffalo" are substantially different. The methodology proposed in this study permits the development of a model capable of answering questions of widely varying degrees of complexity. However, economic considerations as well as input data requirements make this development subject to careful evaluation.

Although classified as "fixed costs", expenditures associated with the other factors, such as the construction of contingency programs and the form of the input data, could be postponed until after a pilot study is completed. Again, these costs depend on the type of "what if" questions asked of the model as well as the data requirements for developing the programs to answer these questions. As previously mentioned, several levels of complexity of contingency programs are foreseen.

More specifically, let us assume that it is possible to delineate the output information requirements of the model; for example, "five classes of age may be sufficient" or "socio-economic background may not be as important as origin". This delineations can be converted into levels of disaggregation (say) from 1 to T where the T^{th} level is the most disaggregated level of desired information. Then

$$FC_i,\ i=1,\ldots,T,$$

can represent the fixed costs of constructing a model where i denotes the level of disaggregation.

The variable costs of running and maintaining the computer model will also depend on the level of disaggregation. More detailed output information may require increased computer tape usage as well as more run time. Moreover, the number of years in the future for which projected information is desired affects the variable costs. The run time for a projection for 1970, three years in the future, is less than for 1977, ten years. Symbolically, these costs can be represented as:

$$(VC_i)_n,\ i=1,\ldots,T,\ n=1,\ldots$$

which states that variable costs not only depend on the level of disaggregation, i, but also on the number of years in the future (n) for which projections are desired.

The final category of costs is that related to the desired precision of the projections. These costs are directly related to the form and content of the data base. The statistical requirements e.g., size of sample, for estimating parameters that yield projections simply of mean values are much less rigorous than those for projections with probabilistically derived confidence intervals. An estimate of "ten thousand engineering students in New York State in 1975" is much less expensive to derive than one of the form: "we are ninety-five percent confident that the actual number of engineering students in 1975 will be

between eleven thousand and nine thousand." Obviously, the latter is more desirable since probability cannot be associated with a single point estimate. Again, the precision requirements must be set by the users of the model, possibly by delineating different levels or categories of precision. Let us assume that the costs of precision are a function of the accessibility of data. A possible categorization might be similar to that described in the preceding section: costs based on (1) subjectively determined estimates; (2) statistical sampling of RECORDS; or (3) use of a higher education data bank. Then

$$(PC_{ij})_n;\ i=1,\ldots,T;\ j=1,\ldots,S;\ n=1,\ldots;$$

can represent the costs of precision which depend on the level of disaggregation, i, the level of precision desired, j, and the number of years in the future of the projection, n.

It should be possible to quantify to some extent the aforementioned costs upon completion of a pilot study. Information on costs of programming, computer run time, sampling activities, input data conversion, etc. can be determined from experience with preliminary runs of the simulation model. Levels of disaggregation and precision can be delineated based on further interviews and discussions with the Advisory Committee.

A more difficult problem is the quantification of the benefits of the computer model. Such benefits as: (1)

a greater understanding of the environment resulting from construction of the model; (2) the ability to analyze the possible effects of complex socio-economic decisions on higher education by using the concept of contingency programs; (3) insight into the complexities of the design and development of an information system; (4) a clear and precise delineation of the enrollment information necessary for higher educational planning in New York State, may not be quantifiable. However, there are two possible categories of benefits that may be amenable to analytical analysis: (1) benefits of more precise projections; (2) benefits realized from greater disaggregation of the projections.

The benefits of a more precise estimate may be approximated by equating them to the costs of deviation between the actual enrollment and its estimate. These costs can be classified into those of overestimating demand and those of underestimating demand. Overestimation costs might include costs of unused educational facilities, while underestimation might incur costs of providing temporary lodgings, classrooms, etc. Note, however, that the cost of not satisfying the demand may be difficult to estimate in such cases as that of inactivity imposed upon creative people — the Professors and research personnel — a condition resulting from overestimation of demand.

Since these benefits depend on the level of precision and the year for which the projection is desired, they can

be represented as follows:

$(PB_j)_n$, j=1,...S, and n=1,... .

The other possible category of benefits is that associated with a greater disaggregation of the projections. Some realizable benefits are (1) more detailed planning for the construction of facilities; (2) better allocation of educational resources; (3) better projection of manpower requirements; and (4) more explicit evaluation of public programs of assistance to higher education. Let these benefits be represented symbolically as

$(BD_i)_n$ i=1,...,T; n=1,...,

showing that the benefits to be derived from disaggregation depend on the level of disaggregation and the number of years in the future of the projection.

These benefits, especially those related to the level of disaggregation, may not be quantifiable now or in the foreseeable future. It is felt, however, that by expressing the quantifiable portion analytically, and letting educational planners evaluate the non-quantifiable aspects, a better evaluation can be made of the worth of the computer model. The following cost/benefits relationship may be useful:

$$(BD_i)_n+(PB_j)_n \geq (VC_i)_n+(PC_{ij})_n+f(FC_i)$$

where $f(FC_i)$ represents a depreciated value for the fixed costs based on the anticipated life of the model. If we can assume that the benefits from greater precision can be

equated to the costs of deviation from the estimates, this relationship can be rewritten to permit the quantifiable portion of the expression to be evaluated against the values that require judgmental decisions by educational planners.

$$(BD_i)_n \geq (VC_i)_n + (PC_{ij})_n + f(FC_i) + (PB_j)_n$$

The above relationship states that the realizable benefits for the computer model must exceed its costs over its useful life. The symbolic representation is not necessary, but does aid in delineating the nature of the costs involved and the possible benefits from constructing a computer model for projecting statewide college enrollments.

The foregoing cost/benefit analysis is presented as a possible approach to evaluation of the successive stages of development of the computer model. Use of this relationship cannot be made until a pilot study involving the programming and testing of a projection model is completed.

APPENDIX A

INSTITUTIONS AND ORGANIZATIONS VISITED DURING THE COURSE OF THIS STUDY

A.1 Colleges and Universities (New York State)

City University of New York

College of Saint Rose

Cornell University

Fordham University

Hofstra University

Hudson Valley Community College

New York University

Rensselaer Polytechnic Institute

State University of New York

Syracuse University

University of Rochester

A.2 Other Institutions and Organizations

American Association of Collegiate Registrars and Admissions Officers

Commissioner of Education, State of Colorado

Institute of Public Affairs, University of California at Los Angeles

State Education Department, New York State

University of Colorado

University of California

U.S. Office of Education, Higher Education Studies

U.S. Office of Education, Program Planning and Evaluation

APPENDIX B

A SIMPLIFIED NUMERICAL EXAMPLE

B.1 Introduction

The material to follow presents a highly simplified numerical example of the operation of the proposed model. The notation utilized will be the same as that of Section 3, and only five years of a simulation run will be presented, i.e., the simplified system will be described beginning with the freshman class of 1969 and will age this class for five years - until 1974. This then will provide a longitudinal analysis of one particular class of students. As is mentioned in Section 3, if cross-sectional analysis of all classes in a particular year were desired, multiple runs would have to be made.

Before proceeding it would be beneficial to review the system represented by the example to follow. First, the freshman class of 1969 will be composed of two principal parts: those who are first-time freshman in 1969 and those, who although they are not first-time freshman in 1969 are, nevertheless freshman in 1969, i.e., they, are in some manner, repeating their freshman year. Each of these will be projected separately, although in an operational model the second part of the freshman class 1969, non-first-time freshman, will have been generated by the model itself.

At this point a (total) raw number of freshman students in 1969 is established. This number is then multiplied by an "initial classification vector" which places students into particular classifications (either colleges or curricula). For instance, in the example which follows, the initial classification vector for freshman in 1969 is (.32, .44, .24). When this vector is multiplied by the projected number of freshman in 1969, it will assign 32% of these freshman to classification one, 44% of them to classification two, etc. Each element of this initial classification vector must also be projected.

Now that freshman students in 1969 are assigned to various classifications (i.e. particular curricula or colleges) we can begin to "age" this population. This is done by multiplying the classification vector for freshman in 1969 by the freshman-sophomore transition matrix for 1969. (Each element of this transition matrix is again projected individually.) The result of this multiplication is to describe what has happened during the 1969-70 school year to those students who were freshman in September 1969. Some of these students will go on as sophomores in September 1970 (either in the curriculum in which they began in 1969 or in a different curriculum), some will leave the educational system, and some will begin again as freshman in 1970. In our numerical example we will concern ourselves only with those students who continue as sophomores in 1970.

The latter students do not constitute the entire sophomore class of 1970. Some sophomores of 1970 were not freshman in 1969. Indeed, some were not even in the educational system in 1969. Therefore to the number of sophomores which has been aged through the model must be added those who (a) were not in the system in 1969, and (b) those who were sophomores in the system in 1969. In this example these will be projected as only <u>one</u> number, but in the operational model, the second part of this number, (b), would automatically be generated by the flows in the simulation. This number is then multiplied by a "sophomore initial classification vector," in our example (.29, .44, .27), which then yields the number of sophomores in 1970, by classification, who were not freshman in 1969. If we add this vector to the sophomores aged through the model we will have the total number of sophomores in 1970, TSOPH(70).

TSOPH(70) is then multiplied by the sophomore-junior transition matrix for 1970 to yield the major portion of juniors in 1971. We then add to this the number of students who will enter the system as juniors in 1971 in the same manner as for sophomores. The same process is then repeated for seniors in 1972 and fifth-year students in 1973.

Characteristics of students are provided by utilizing "output breakdown vectors". Each particular level (freshman, sophomore, etc.), year (1970, 1971, etc.,) and classification

(engineering, liberal arts, etc.) will have a potentially unique output breakdown vector. For example, our model will be shown to project 8,814 fifth-year pure-science students in 1973. A potential output breakdown vector for fifth-year, pure-science students in 1973 could be:

	Age		Geographical Origin		
% Male	$\leq$ 23	$\geq$ 23	1	2	3
(.79	.76	.24	.37	.19	.44)

By multiplying 8,814 by this vector we would find that 6,465 of these students were male, 6,220 were 23 years of age or less, 3,261 were from geographical origin 1, etc. Again every element of each output breakdown vector constitutes a distinct projection based on a longitudinal analysis of data.

B.2 The Example

The purpose of this section is to present a detailed numerical analysis of the flows in the proposed model. Although realistic numbers are used in this example it should not be inferred that the structure of the methodology has been finalized prior to the completion of a pilot study. Rather, it is hoped that the presentation of this numerical example will aid in the general conceptualization of the presently proposed model. It is quite probable that as research on the model progresses many of the currently envisaged constructs will be modified and adapted to ensuing

developments. However, the basic theme of a dynamic flow model should remain invariant.

The following simplifying assumptions have been employed throughout this example.

(i) In the actual model, projections would have to be made for first time freshman, sophomore transfers, etc., and each element of every transition matrix, initial classification vector and output breakdown vector. These problems are not mentioned further; the material to follow will focus on the "flow nature of the proposed methodology".

(ii) Only three classifications of college or curricula are considered: (1) science-engineering curricula; (2) social science and education curricula, and (3) liberal arts curricula. These are assumed to be mutually exclusive and collectively exhaustive.

(iii) A six element output breakdown vector is employed. This permits six classifications of student characteristics: sex, 2 age groups, and 3 geographical origins.

The numerical operations themselves are presented diagrammatically in order that the clarity of the flow concept can be maintained. The deletion of the numerical details (such as actual multiplications, additions, etc.)

should then make comprehension less difficult. The diagrams themselves follow the schema presented in the introduction to this section.

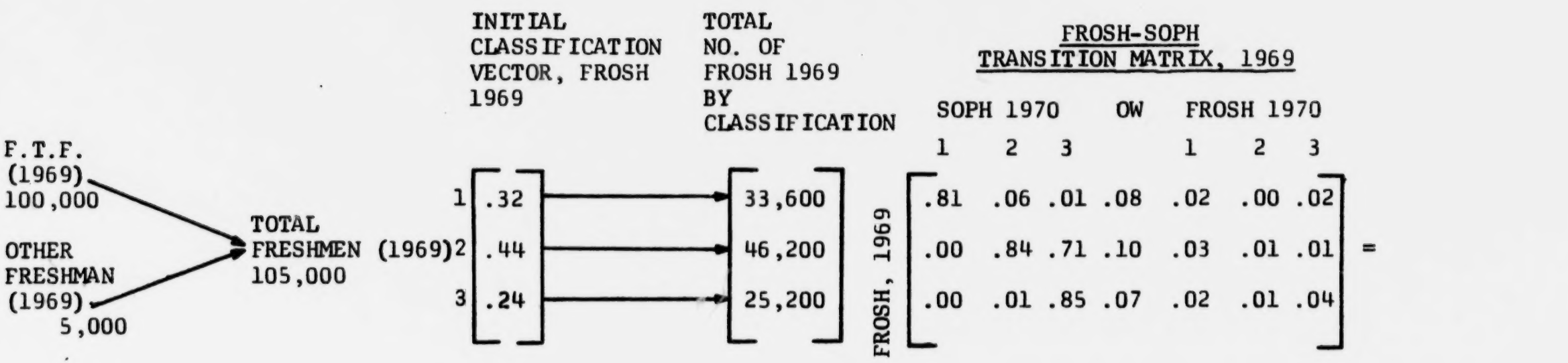
B.3 Summary

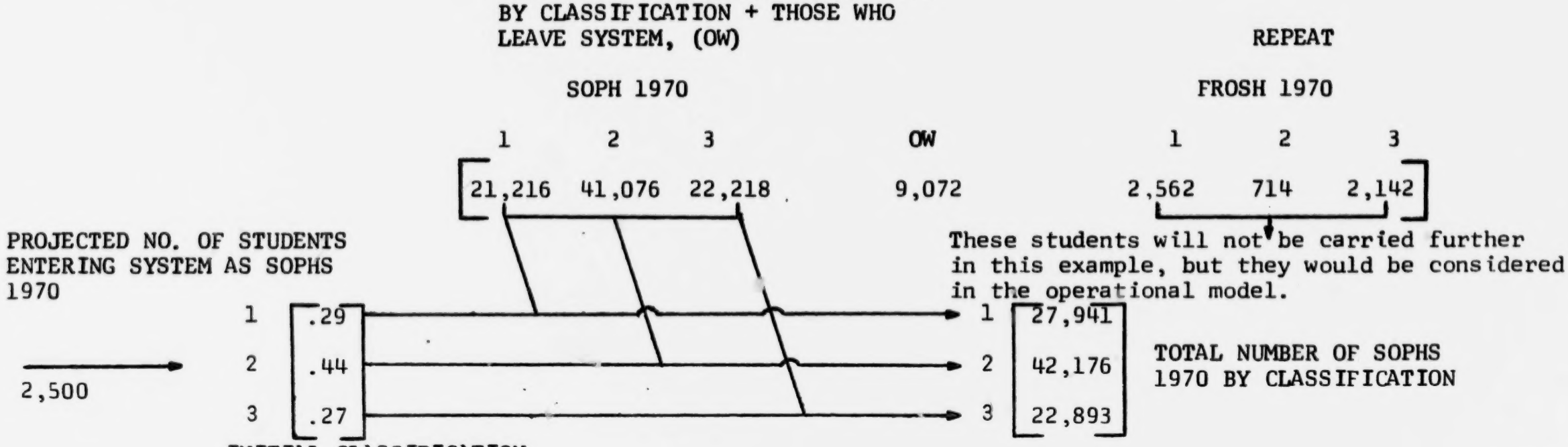
It should be re-emphasized that this example is highly simplified. A major portion of the actual projections used, those of model parameters, have been assumed as "given". In fact each of the ratios used throughout this example requires a separate projection.

Secondly, the scope of this example is considerably more narrow than that of an operational model. Thus the model allows for more than three classifications (curriculum or college) and more than six characteristics of output.

Finally, only a single class of students, freshman 1969, has been aged through the model. As is stated in Section 3, comparable agings would be processed if a broader cross-sectional categorization were required.

The scope of the operational model consequently dictates more complexity than can be demonstrated in this simple example. On the other hand the potential benefits of flexibility and adapability designed into the proposed model more than offset its inherent complexity.


INITIAL CLASSIFICATION VECTOR, FROSH 1969
TOTAL NO. OF FROSH 1969 BY CLASSIFICATION
FROSH-SOPH TRANSITION MATRIX, 1969
SOPH 1970
OW
FROSH 1970
1 2 3
1 2 3
F.T.F. (1969) 100,000
OTHER FRESHMAN (1969) 5,000
TOTAL FRESHMEN (1969) 105,000
1 .32
2 .44
3 .24
33,600
46,200
25,200
FROSH, 1969
.81 .06 .01 .08 .02 .00 .02
.00 .84 .71 .10 .03 .01 .01
.00 .01 .85 .07 .02 .01 .04
=


SOPH 1970 & REPEATING FROSH 1970 BY CLASSIFICATION + THOSE WHO LEAVE SYSTEM, (OW)
REPEAT FROSH 1970
SOPH 1970
1 2 3
OW
21,216 41,076 22,218
9,072
1 2 3
2,562 714 2,142
These students will not be carried further in this example, but they would be considered in the operational model.
PROJECTED NO. OF STUDENTS ENTERING SYSTEM AS SOPHS 1970
2,500
1 .29
2 .44
3 .27
INITIAL CLASSIFICATION VECTOR, SOPH 1970
1 27,941
2 42,176
3 22,893
TOTAL NUMBER OF SOPHS 1970 BY CLASSIFICATION

TOTAL NUMBER OF SOPHS, 1970
BY CLASSIFICATION

1	27,941
2	42,176
3	22,893

SOPH-JUNIOR TRANSITION MATRIX, 1970

SOPH, 1970	Junior 1971 1	2	3	OW	Soph. 1971 1	2	3
1	.80	.04	.00	.11	.04	.01	.00
2	.00	.82	.00	.09	.04	.05	.00
3	.00	.00	87	.07	.01	.01	.04

=

Juniors, 1971 and repeat Sophs 1971 by classification and those who leave system, (OW)

Juniors 1971 1	2	3	OW	Soph 1971 1	2	3
22,353	35,702	19,917	8,472	2,464	2,617	916

PROJECTED NO. OF STUDENTS ENTERING SYSTEM AS JUNIORS 1971

9,000 →

1	.27
2	.45
3	.28

INITIAL CLASSIFICATION VECTOR, JUNIORS 1971

TOTAL NUMBER OF JUNIORS 1971 BY CLASSIFICATION

1	24,783
2	39,752
3	22,437

TOTAL NUMBER OF JUNIORS 1971
BY CLASSIFICATION

1	24,783
2	39,752
3	22,437

JUNIORS, 1971

JUNIOR-SENIOR TRANSITION MATRIX, 1971

	Seniors 1972			OW	Juniors 1972		
	1	2	3		1	2	3
1	.86	.01	.01	.06	.05	.01	.00
2	.00	.88	.00	.07	.00	.05	.00
3	.00	.00	.94	.05	.00	.00	.01

=

SENIORS 1972 AND REPEAT JUNIORS 1972 BY CLASSIFICATION AND THOSE WHO LEAVE SYSTEM

Seniors 1972			OW	Juniors 1972		
1	2	3		1	2	3
21,313	35,380	21,315	5,392	1,239	2,235	224

PROJECTED NUMBER OF STUDENTS ENTERING SYSTEM AS SENIORS 1972

1	.25
2	.44
3	.31

410 →

INITIAL CLASSIFICATION VECTOR, SENIORS 1972

TOTAL NUMBER OF SENIORS, 1972 BY CLASSIFICATION

1	21,416
2	35,560
3	21,442

TOTAL NUMBER OF SENIORS 1972 BY CLASSIFICATION

1	21,416
2	35,560
3	21,442

SENIOR-FIFTH-YEAR TRANSITION MATRIX, 1972

SENIORS, 1972	FIFTH 1973 1	2	3	OW	SENIORS 1973 1	2	3
1	.31	.12	.00	.52	.05	.00	.00
2	.00	.26	.00	.69	.00	.05	.00
3	.00	.08	.27	.64	.00	.00	.01

=

FIFTH-YEAR STUDENTS, 1973 AND REPEAT SENIORS 1973 BY CLASSIFICATION AND THOSE WHO LEAVE SYSTEM

FIFTH 1973 1	2	3	OW	SENIORS 1973 1	2	3
6,639	13,531	5,789	48,395	1,071	1,778	214

In the operational model these would be further broken down by the number of seniors receiving degrees in various fields. Likewise, fifth-year students would also be further categorized by the numbers receiving degrees in various classifications

Projected number of students entering system as fifth-year students, 1973

1	.29
2	.42
3	.29

7,500 →

INITIAL CLASSIFICATION VECTOR, FIFTH-YEAR 1973

Total number of fifth-year students, 1973 by classification

1	8,814
2	16,681
3	7,964

APPENDIX C

THE DESIGN OF A SAMPLING PLAN FOR ESTIMATING STUDENT FLOWS

The purpose of this Appendix is to describe in detail a feasible method of statistically sampling to estimate student flows. In addition, this approach could be used to develop estimates of student characteristics such as origin, sex, and marital status.

As discussed in Section 2 the magnitudes of the numbers of curricula and colleges in New York State require, for a manageable model, some classification. This classification would have to be made by those responsible for higher educational planning in New York State, i.e., the "users" of the simulation model. One possible approach is to classify colleges and curricula by: (1) sources of funds, public or private, (2) type of institution, complex university, four year college or two-year college, and (3) curricula - science or non-science. This yields twelve possible classifications of students, e.g., public-complex university-science, private-complex university-non-science, etc. [It might be noted that the model described in Section 3 could operate using the classifications.]

Once this classification has been made, all classes will be assessed with respect to their susceptibility to transfers. Information obtained from completed questionnaires

by or personal interviews with knowledgeable individuals at the respective colleges will be useful in making the above assessment. In fact, initial (subjective) transition probabilities should be available from such a survey. This information would also aid in arriving at suggested sizes of samples of student records from various colleges and curricula at selected institutions from the various classes.

If, for example, a registrar at a public, complex university states that about 20% of his second-year science students transfer in from the same curriculum in a private, complex university after their freshman year, then based on this estimate, approximately 1540 student records must be sampled to have ninety-five percent chance that our estimate of the true proportion that make this transfer lies within two percent of the true value.

Figure 9 gives approximate sample sizes for different levels of confidence and precision for this type of a sampling plan. Note that as more precision is required and/or a greater degree of confidence in the estimate, then larger sample sizes are required.

As discussed in Section 4.2, the above approach is technically feasible. The question that must be answered is whether the value of probabilistic estimates as opposed to subjective evaluations is worth the additional incremental cost. Naturally, the implementation of a statewide data system would eventually alleviate sampling difficulties.

SAMPLE SIZES TO ESTIMATE p WITH INDICATED PRECISION AND CONFIDENCE LEVELS

p \ LEVEL OF CONFIDENCE	Precision of ± 0.2p			Precision of ± 0.1p		
	.80	.90	.95	.80	.90	.95
.1	370	609	864	1479	2435	3457
.2	164	271	384	657	1082	1537
.3	96	158	224	383	631	896
.4	62	101	144	247	406	576
.5	41	68	96	164	271	384

p \ LEVEL OF CONFIDENCE	Precision of ± .06			Precision of ± .04			Precision of ± .02		
	.80	.90	.95	.80	.90	.95	.80	.90	.95
.1	41	68	96	92	152	216	370	609	864
.2	73	120	171	164	271	384	657	1082	1537
.3	96	158	224	216	355	504	863	1421	2016
.4	110	180	256	247	406	576	986	1624	2305
.5	114	188	267	257	423	600	1027	1691	2401

Figure 13

APPENDIX D

COMPUTER MODEL ADVISORY COMMITTEE

Donald Axelrod - Assistant Director - Division of Budget

Lester Brookner - Director, Institutional Research & Planning - New York University

James Fitzgibbons - President - Hudson Valley Community College

Edward Hollander - Coordinator of Master Plan - City University

Francis Horn - President - Commission of Independent Colleges

Les Ingalls - Executive Secretary - Ass. of Colleges & Universities of New York State

Arthur Kaiser - Admissions Office - State University of New York at Buffalo

George Knerr - Dean for Students Personnel - Pace College

Lionel Livesey - Vice President - State University of New York

Harold Martin - Chancellor - Union University

Tom Mason - Director of Planning - Rochester University

Harry Mills - Vice President & Provost - St. John's University

Edwin Smith - Registrar - Syracuse University

Ralph VonGuerard - Registrar - New York University